

# **Advances in Solid State Physics**

Bernhard Kramer (Ed.)

**44**



**Springer**

# **Advances in Solid State Physics**

## **Volume 44**

Available online at  
**SpringerLink.com**

Advances in Solid State Physics is part of [SpringerLink] service. For all customers with standing orders for Advances in Solid State Physics we offer the full text in electronic form via [SpringerLink] free of charge. Please contact your librarian who can receive a password for free access to the full articles by registration at:

[springerlink.com](http://springerlink.com) → Orders

If you do not have a standing order you can nevertheless browse through the table of contents of the volumes and the abstracts of each article at:

[springerlink.com](http://springerlink.com) → Browse Publications

There you will also find more information about the series.

# Advances in Solid State Physics

---

Advances in Solid State Physics is a book series with a history of about 50 years. It contains the invited lectures presented at the Spring Meetings of the “Arbeitskreis Festkörperphysik” of the “Deutsche Physikalische Gesellschaft”, held in March of each year. The invited talks are intended to reflect the most recent achievements of researchers working in the field both in Germany and worldwide. Thus the volume of the series represent a continuous documentation of most recent developments in what can be considered as one of the most important and active fields of modern physics. Since the majority of invited talks are usually given by young researchers at the start of their career, the articles can also be considered as indicating important future developments.

The speakers of the invited lectures and of the symposia are asked to contribute to the yearly volumes with the written version of their lecture at the forthcoming Spring Meeting of the Deutsche Physikalische Gesellschaft by the Series Editor early before the meeting. Articles are submitted during the meeting. All articles are refereed.

Advances in Solid State Physics is addressed to all scientists at universities and in industry who wish to obtain an overview and to keep informed on the latest developments in solid state physics. The language of publication is English.

## Series Editor

Prof. Bernhard Kramer

I. Institut für Theoretische Physik

Universität Hamburg

Jungiusstraße 9

20355 Hamburg

Germany

[kramer@physnet.uni-hamburg.de](mailto:kramer@physnet.uni-hamburg.de)

Bernhard Kramer (Ed.)

# Advances in Solid State Physics

44

With 267 Figures

**Prof. Bernhard Kramer (Ed.)**  
I. Institut für Theoretische Physik  
Universität Hamburg  
Jungiusstraße 9  
20355 Hamburg  
Germany  
[kramer@physnet.uni-hamburg.de](mailto:kramer@physnet.uni-hamburg.de)

**Physics and Astronomy Classification Scheme (PACS): 60.00; 70.00; 80.00**

**ISSN print edition: 1438-4329**  
**ISSN electronic edition: 1617-5034**

**ISBN 978-3-540-21148-8   ISBN 978-3-540-39970-4**  
**DOI 10.1007/978-3-540-39970-4**

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

[springeronline.com](http://springeronline.com)

© Springer-Verlag Berlin Heidelberg 2004

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting by the authors using a Springer TeX macro package  
Final processing: DA-Tex · Gerd Blumenstein · [www.da-tex.de](http://www.da-tex.de)  
Production: LE-Tex GbR, Leipzig, [www.le-tex.de](http://www.le-tex.de)

Cover concept using a background picture by Dr. Ralf Stannarius, Faculty of Physics and Earth Sciences, Institute of Experimental Physics I, University of Leipzig, Germany  
Cover design: *design & production* GmbH, Heidelberg

# Preface

The 2004 Spring Meeting of the Arbeitskreis Festkörperphysik of the Deutsche Physikalische Gesellschaft was held during March 8 to 12, 2004 in Regensburg, one of our traditional meeting places. The number of conference attendees this time was again at a record high: about 3700, mostly young physicist, participated in the largest of the spring meetings of the DPG and the second largest worldwide. This confirms that Solid State Physics remains a field of continuously increasing attractivity which is of outstanding importance especially for applications.

This time, the total number of conference contributions was close to 3000, including posters, short oral contributions, invited talks and plenary lectures. The present volume 44 of the “Advances in Solid State Physics” contains the written versions of most of the invited and plenary talks. Several of the topical talks given at the numerous, very lively symposia are also included. Most of these were organized collaboratively by several divisions. These covered extremely interesting and timely subjects, and attracted much interest, especially among the young participants.

Hamburg, April 2004

*Bernhard Kramer*

# Contents

---

## Part I Solar Cells

---

### Slow Solar Ascent

<i>Hans J. Queisser</i> .....	3
1 Introduction: Junctions .....	3
2 Silicon Conquest .....	4
3 A Novel Solar Energy Converter .....	5
4 Signals from Space .....	7
4.1 Early Interpretation .....	8
4.2 Nonideal Junctions .....	8
4.3 Silicon Valley .....	9
4.4 Detailed Balance .....	9
5 Third Generation .....	10
References .....	11

### Interface Analysis of Thin Film Solar Cells

#### Using Photoelectron Spectroscopy

<i>Andreas Klein</i> .....	13
1 Interfaces in Thin Film Solar Cells .....	13
2 Band Alignment .....	14
3 Photoelectron Spectroscopy .....	16
4 Experimental Procedure .....	16
5 Example Results .....	18
5.1 CdTe / Metal Interfaces .....	18
5.2 CIGS / II-VI Compound Interfaces .....	20
5.3 TCO Interfaces .....	21
6 Conclusions .....	22
References .....	23

### Electronic Properties of Cu(In,Ga)Se<sub>2</sub> Thin-Film Solar Cells

#### – An Update

<i>Uwe Rau</i> .....	27
1 Introduction .....	27
2 Device Technology .....	28
2.1 Absorber Preparation and the Role of Na .....	28

## VIII Contents

2.2	Buffer and Window Deposition . . . . .	29
3	Device Physics . . . . .	30
3.1	Short Circuit Current . . . . .	30
3.2	Open Circuit Voltage . . . . .	31
3.3	Surface Band Gap Widening . . . . .	32
3.4	Defects . . . . .	34
3.5	Polycrystallinity . . . . .	35
4	Conclusions . . . . .	36
	References . . . . .	36

### **Progress in Manufacturable High-Efficiency Silicon Solar Cells**

<i>Rudolf Hezel</i> . . . . .	39	
1	Introduction . . . . .	39
2	Silicon Substrate Options . . . . .	40
3	High Efficiency Silicon Solar Cells . . . . .	42
3.1	The Point-Contact Solar Cell . . . . .	43
3.2	The HIT Solar Cell . . . . .	43
3.3	The “Buried Contact” Solar Cell . . . . .	44
3.4	The Standard OECO Solar Cell . . . . .	45
3.5	Back Contacted Bifacial “BACK OECO” Solar Cell . . . . .	46
4	Conclusion . . . . .	47
	References . . . . .	48

### **Second and Third Generation Photovoltaics – Dreams and Reality**

<i>Jürgen H. Werner</i> . . . . .	51	
1	Introduction . . . . .	51
2	Learning Curve and Module Market . . . . .	52
2.1	Learning Curve of Crystalline Silicon Modules . . . . .	52
2.2	Market Growth . . . . .	53
3	Electricity Cost . . . . .	53
3.1	Dependence on Insolation . . . . .	53
3.2	Cost Degradation . . . . .	54
3.3	Areas and Production Rate of PV Modules . . . . .	55
4	Second Generation Hopes: Thin-Film Materials . . . . .	56
4.1	Amorphous Si (a-Si) . . . . .	58
4.2	CdTe . . . . .	58
4.3	Cu(In,Ga)Se <sub>2</sub> (CIGS) . . . . .	59
4.4	Crystalline Si (c-Si) . . . . .	60
5	Third Generation Dreams: New Physics . . . . .	62
6	Conclusions . . . . .	63
	References . . . . .	64

---

**Part II Optical Properties**


---

<b>Optical Transmission</b>	
<b>through Periodically Nano-structured Metal Films</b>	
<i>L. Martín-Moreno, F. J. García-Vidal</i> .....	69
1 Introduction .....	69
2 SPs as a Route to Sub-wavelength Optics .....	70
3 Optical Transmission	
Through Nano-structured Metals .....	71
3.1 1D Slit Array .....	71
3.1.1 Theoretical Formalism .....	72
3.1.2 Results .....	74
3.2 2D Hole Array .....	76
3.2.1 Minimal Model .....	77
3.3 Single Apertures Flanked by Corrugations .....	79
References .....	79
<b>Charged Excitons in the Quantum Hall Regime:</b>	
<b>Optical Probe of Fractionally Charged Quasiholes</b>	
<i>Christian Schüller, Kay-Birger Broocks, Patrick Schröter, Christian Heyn, Detlef Heitmann, Max Bichler, Werner Wegscheider, Vladimir M. Apalkov, Tapash Chakraborty</i> .....	81
1 Introduction .....	81
2 Magneto-photoluminescence and Absorption	
in the Quantum Hall Regime .....	83
2.1 Experimental Details .....	83
2.2 High Electron Density .....	83
2.3 Intermediate and Low Electron Density .....	85
3 Theoretical Model:	
Creation of Fractionally-Charged Quasiholes .....	88
4 Conclusion .....	91
References .....	91
<b>Three-Dimensional Lithography of Photonic Crystals</b>	
<i>A. Blanco, K. Busch, M. Deubel, C. Enkrich, G. von Freymann, M. Hermatschweiler, W. Koch, S. Linden, D. C. Meisel, G. A. Ozin, S. Pereira, C. M. Soukoulis, N. Tétreault, M. Wegener</i> .....	93
1 Introduction .....	93
2 Holographic Lithography .....	94
2.1 The Photoresist SU-8 .....	94
2.2 The Crystallography of Multiple-beam Interference Patterns ..	95
3 Direct Laser Writing .....	96
3.1 Multi-photon Polymerization .....	97
3.2 Experimental Realization .....	97

3.3	Direct Laser Writing of Three-dimensional Photonic Crystals . . . . .	98
3.3.1	Woodpile Photonic Crystals . . . . .	98
3.3.2	SP <sub>2</sub> Photonic Crystals . . . . .	100
4	Outlook: Inversion . . . . .	101
	References . . . . .	102

**Photonic Crystal Waveguides:****Dispersion, Anomalous Refraction and Applications**

<i>Remigius Zengerle</i> . . . . .	105	
1	Introduction . . . . .	105
2	One- and Two- Dimensional Periodic Waveguides . . . . .	106
3	Anomalous Refraction Phenomena in One-Dimensional Periodic Waveguides . . . . .	106
3.1	Simultaneous Positive and Negative Ray Refraction . . . . .	106
4	Refraction in Two-Dimensional Planar Photonic-Crystal Waveguides . . . . .	107
4.1	Frequency Dependent Propagation in 2D Photonic Crystal Waveguides . . . . .	108
4.2	Anomalous Super Refraction . . . . .	109
4.3	Photonic Bandgaps . . . . .	109
4.4	Normal Super Refraction . . . . .	111
4.5	Generalized Refraction Law . . . . .	111
4.6	Interference of Floquet-Bloch Waves Excited by a Gaussian Beam . . . . .	112
4.7	Negative Ray Refraction in State-of-the-Art III-V Semiconductor Waveguides . . . . .	113
5	Applications . . . . .	113
5.1	Frequency-Dependent Spot-Size Transformation . . . . .	113
5.2	Spot-Size Transformation in State-of-the-Art III-V Semiconductor Waveguides . . . . .	114
5.3	Further Applications . . . . .	115
6	Conclusions . . . . .	115
	References . . . . .	116

**Photonic Crystal Structures in Ultrafast Optics**

<i>A. Tünnermann, T. Schreiber, M. Augustin, J. Limpert, M. Will, S. Nolte, H. Zellmer, R. Iliew, U. Peschel, F. Lederer</i> . . . . .	117	
1	Introduction . . . . .	117
2	Two-Dimensional Photonic Crystal Slab Waveguides . . . . .	119
2.1	Filtering Devices Based on Photonic Crystal Waveguides . . . . .	122
2.2	Photonic Crystal Waveguide Bends . . . . .	123
3	Air-Core Photonic Band Gap Fibers . . . . .	125
3.1	Properties of an Air-Core PCF . . . . .	126
3.2	Megawatt Peak Power fs Pulse Delivery through Air-Core PCFs . . . . .	128

3.3 All Fiber CPA System Using Compression in an Air-Core PCF .....	129
4 Conclusions and Outlook .....	130
References .....	131

### Part III Quantum Transport

#### Novel Non-equilibrium Zero-resistance States in the High Mobility GaAs/AlGaAs System

<i>R. G. Mani</i> .....	135
1 Introduction .....	135
2 Experiment .....	135
3 Results .....	136
References .....	145

#### Oscillatory Photoconductivity

#### of a Two-Dimensional Electron Gas in Magnetic Field

<i>I.A. Dmitriev, M.G. Vavilov, I.L. Aleiner, A.D. Mirlin, D.G. Polyakov</i> .....	147
1 Introduction .....	147
2 Non-equilibrium Distribution Function Induced by Microwave Radiation .....	148
3 Oscillatory Photoconductivity .....	150
4 Strong Magnetic Fields: Separated Landau Levels .....	153
5 Temperature Dependence: Inelastic Relaxation Rate .....	154
6 Summary and Comparison with Experiment .....	155
7 Conclusions .....	155
References .....	156

#### Molecular Wires in Electromagnetic Fields

<i>Sigmund Kohler, Jörg Lehmann, Michael Strass, Peter Hänggi</i> .....	157
1 The Wire-Lead Model .....	158
2 Floquet Transport Theory .....	159
3 Conduction Properties of Driven Molecular Bridges .....	161
3.1 Resonant Excitations .....	162
3.2 Numerical Results .....	164
4 Conclusions .....	164
References .....	166

#### Coulomb Drag as a Probe of Quantum Hall Systems

<i>Felix von Oppen, I.V. Gornyi, A.D. Mirlin</i> .....	169
1 Coulomb Drag in High Landau Levels .....	170
2 Previous Theoretical Work .....	171
3 Diagrammatic Theory .....	172
3.1 The Triangle Vertex .....	173
3.2 Screened Interlayer Interaction .....	175

3.3 Results for the Drag Resistivity .....	175
4 Summary and Outlook .....	177
References .....	178

---

## Part IV Quantum Dots, Wires and Rings

---

### Applications of Adiabatic Passage in Solid-State Devices

<i>Jens Siewert, Tobias Brandes</i> .....	181
1 Introduction .....	181
2 Adiabatic Passage in Three-Level Atoms .....	182
3 Analogy for a Cooper-Pair Box with Two Islands .....	184
4 Solid-State Applications of Adiabatic Passage .....	185
4.1 Non-abelian Holonomies by Sequences of Adiabatic Population Transfers .....	186
4.2 Coupled Quantum Dots .....	187
5 Conclusions .....	188
References .....	189

### Quantum Dots: Building Blocks of Quantum Devices?

<i>Manfred Bayer, G. Ortner, A. Larionov, D.R. Yakovlev, M. Schwab, P. Borri, W. Langbein, U. Woggon, I. Yugova, G. Baldassarri Höger von Högersthal, Y.B. Lyanda-Geller, T.L. Reinecke, S. Fafard, Z. Wasilewski, M. Korkusinski, P. Hawrylak, A. Forchel, J.P. Reithmaier</i> .....	191
1 Introduction .....	191
2 Coupled Quantum Dot Structures .....	194
3 Exciton States in Quantum Dot Molecules .....	195
4 Proof of Coherent Dot Coupling .....	198
5 Electric Field Control of Molecule Bonding .....	205
6 Exciton Dephasing .....	207
7 Summary .....	210
References .....	210

### The Structure and Dispersion of a Sharp Quantum Hall Edge Probed by Momentum-Resolved Tunneling

<i>M. Huber, M. Grayson, M. Rother, W. Biberacher, W. Wegscheider, G. Abstreiter</i> .....	213
1 Introduction .....	213
2 Sample Design .....	215
3 $V = 0$ : structure of a Sharp Edge .....	215
3.1 The Momentum Conservation Picture: $k_y$ .....	216
3.2 The Orbit-Center Conservation Picture: $X$ .....	217
3.3 Real-Space Position of Sharp Edge Channels .....	218
3.4 Lineshape Analysis .....	219
3.5 Substructure beyond Outermost Edge .....	220

3.6 Width of Conductance Peak .....	220
4 $V < 0$ : Dispersion of a Sharp Edge .....	221
5 Conclusion .....	223
References .....	224

## **Structural Properties of Semiconductor Nanostructures from X-Ray Scattering**

<i>J. Stangl, T. Schülli, A. Hesse, V. Holy, G. Bauer, M. Stoffel, O.G. Schmidt</i> .....	227
1 Introduction .....	227
2 Coplanar X-Ray Diffraction .....	228
3 Iso-Strain Scattering .....	229
3.1 Interdependence of Strain, Lateral Size, and Height .....	230
3.2 Composition from Anomalous Scattering .....	231
4 Experiment .....	231
4.1 Influence of Ge Deposition Temperature .....	232
4.2 Influence of Capping Temperature .....	234
4.3 Strain Distribution .....	235
5 Summary .....	236
References .....	236

## **Part V Quantum Correlations beyond the Fermi Liquid**

### **Spectroscopic Signatures of Spin-Charge Separation in a Quasi-One-Dimensional Organic Conductor**

<i>Ralph Claessen, Michael Sing</i> .....	241
1 Introduction .....	241
2 Organic Charge-Transfer Salts as Strongly Correlated Quasi-1D Metals .....	242
3 Experimental Details .....	243
4 Results and Discussion .....	244
5 Conclusions .....	250
References .....	250

### **Quantum Critical Metals: Beyond the Order Parameter Fluctuations**

<i>Qimiao Si</i> .....	253
1 The Order Parameter Fluctuation Theory of Quantum Critical Points and Its Breakdown .....	253
2 Quantum Critical Heavy Fermions .....	255
3 Microscopic Results and Their Robustness .....	256
3.1 Microscopic Results .....	256
3.2 Beyond Microscopics .....	258
3.3 In What Sense Is the QCP Local? .....	258
4 Experiments .....	259

## XIV    Contents

4.1	Spin Dynamics . . . . .	259
4.2	Thermodynamics . . . . .	260
4.3	Electronic Measurements . . . . .	261
5	Broader Context . . . . .	261
6	Summary . . . . .	262
	References . . . . .	263

### **Quantum Monte Carlo Simulation of Confined Bosonic Atoms in Optical Lattices**

<i>Stefan Wessel, Fabien Alet, Matthias Troyer, George Batrouni</i> . . . . .	265	
1	Introduction . . . . .	265
2	Model and Numerical Techniques . . . . .	266
3	Simulations of Realistic 2D Traps . . . . .	267
3.1	Phase Coexistence in Trapped Systems . . . . .	267
3.2	Local Compressibility . . . . .	268
3.3	Local Potential Approximation . . . . .	269
4	Effective Ladder Model for the Superfluid Layers . . . . .	270
4.1	The Effective Ladder Model . . . . .	270
4.2	Results for the Ladder Model Compared with a Realistic Trap . . . . .	271
4.3	Absence of Quantum Critical Behavior . . . . .	272
5	Discussion and Conclusion . . . . .	273
	References . . . . .	274

### **Cold Atoms in Optical Lattices**

<i>Wilhelm Zwerger</i> . . . . .	277	
1	Introduction . . . . .	277
2	Mott-Hubbard Transition for Bosons . . . . .	280
3	Atomic Quantum Wires . . . . .	287
4	Perspectives . . . . .	292
	References . . . . .	293

---

## **Part VI    Thin Films and Surfaces**

---

### **Plasma-Assisted Deposition and Crystal Growth of Thin Indium-Tin-Oxide (ITO) Films**

<i>R. Hippler, H. Steffen, M. Quaas, T. Röwf, T.M. Tun, H. Wulf</i> . . . . .	299	
1	Introduction . . . . .	299
2	Experiment . . . . .	301
3	Results . . . . .	302
3.1	Chemical Composition Determined by Photoelectron Spectroscopy (XPS) . . . . .	303
3.1.1	Oxygen Dependence . . . . .	303
3.1.2	Bias Voltage Dependence . . . . .	303
3.2	Film Properties Determined by GIXD . . . . .	304

3.3	Post-annealing of Deposited Indium-Tin-Oxide Films . . . . .	305
3.4	Deposition Rate and Film Density Determined by GIXR . . . . .	306
3.5	Oxidized Target . . . . .	306
3.6	Ion Energy Distribution Function (IEDF) . . . . .	307
3.7	Energy Influx . . . . .	309
4	Summary . . . . .	311
	References . . . . .	311

## Strains and Stresses in GaN Heteroepitaxy – Sources and Control

*A. Dadgar, R. Clos, G. Strassburger, F. Schulze, P. Veit, T. Hempel, J. Bläsing, A. Krtschil, I. Daumiller, M. Kunze, A. Kaluza, A. Modlich, M. Kamp, A. Diez, J. Christen, A. Krost . . . . .* 313

1	Experimental . . . . .	313
2	Sources of Stresses and Strains . . . . .	314
2.1	Seed Layer Growth . . . . .	314
2.2	Lattice Mismatch . . . . .	316
2.3	Doping . . . . .	316
2.4	Thermal Stress . . . . .	318
3	Strain Engineering . . . . .	319
3.1	Thermal Stress . . . . .	319
3.2	Intrinsic Stress . . . . .	319
3.3	Doping . . . . .	320
4	Devices . . . . .	321
4.1	Transistors . . . . .	322
4.2	LEDs . . . . .	322
5	Summary . . . . .	323
	References . . . . .	323

## Morphological Transition in Thin Lamellar Diblock Copolymer Films as Revealed by Combined GISAXS and AFM Studies

*Christine M. Papadakis, Peter Busch, Dorthe Posselt, Detlef-M. Smilgies . . . . .* 327

1	Introduction . . . . .	327
2	Methods Used for Investigating Thin Lamellar Diblock Copolymer Films . . . . .	328
3	The Lamellar Orientation in Thin Poly(styrene- <i>b</i> -butadiene) Films . . . . .	330
4	Discussion . . . . .	334
	References . . . . .	336

### **III-V Semiconductors Under Ion Bombardment – Studied by RBS in situ at 15 K**

<i>E. Wendler, W. Wesch</i> .....	339
1 Introduction .....	339
2 Experimental Methods .....	341
3 Mechanisms of Damage Formation .....	341
3.1 Type A Materials – Example GaAs .....	342
3.2 Type B Materials – Example GaN .....	343
3.3 Effect of Ion Species .....	345
4 Systematization of Damage Production for all III-V Compounds Investigated .....	347
5 Summary .....	349
References .....	350

### **Size Controlled Si Nanocrystals**

<i>Margit Zacharias</i> .....	351
1 Introduction .....	351
2 Experimental .....	353
2.1 Sample Preparation .....	353
2.2 Measurements .....	353
3 Results .....	354
3.1 Size, Density and Distribution Control .....	354
4 Discussion .....	357
4.1 Si Nanocrystals .....	357
4.2 Er Doping .....	357
5 Conclusion .....	358
References .....	359

## **Part VII Disordered Systems and Soft Matter**

### **Fat Tail Statistics and Beyond**

<i>Joachim Peinke, Malte Siefert, Stephan Barth, Christoph Renner, Falk Riess, Matthias Wächter, Rudolf Friedrich</i> .....	363
1 Introduction .....	363
2 Turbulence .....	364
3 Finance .....	366
4 Surface Roughness .....	367
5 Stochastic Process for Scale Dependent Complexity .....	368
6 Discussion and Conclusion .....	371
References .....	372

<b>Statistical Physics for Complex Cosmic Structures</b>	
<i>Luciano Pietronero, Francesco Sylos Labini</i> . . . . .	375
1 Introduction . . . . .	375
2 Galaxy Structures . . . . .	375
3 Primordial Density Fields . . . . .	380
4 The Problem of Gravitational Structure Formation . . . . .	383
5 Conclusion . . . . .	386
References . . . . .	386
<b>Local Scale Invariance in Ageing Phenomena</b>	
<i>Malte Henkel</i> . . . . .	389
1 Introduction . . . . .	389
2 Phenomenology of Ageing . . . . .	389
3 Local Scale Invariance . . . . .	392
4 Numerical Test in the Glauber-Ising Model . . . . .	394
5 Influence of Noise . . . . .	397
References . . . . .	400
<b>Phase Transitions and Segregation Phenomena in Vibrated Granular Systems</b>	
<i>Christof A. Kruelle, Sébastien Aumaître, Andreas P. J. Breu, Andreas Goetzendorfer, Tobias Schnautz, Rafał Grochowski, Peter Walzel</i> . . . . .	401
1 Introduction . . . . .	401
2 Horizontal Circular Vibration . . . . .	402
3 Vertical Vibration . . . . .	407
4 Vertical Circular Vibration . . . . .	410
5 Conclusion . . . . .	412
References . . . . .	412
<b>Ab-Initio Thermodynamics of Metal Alloys: From the Atomic to the Mesoscopic Scale</b>	
<i>Stefan Müller</i> . . . . .	415
1 Introduction . . . . .	415
2 Methods . . . . .	416
3 Precipitation . . . . .	418
4 Short-Range Order and Mixing Enthalpies . . . . .	422
5 Summary . . . . .	424
References . . . . .	425
<b>Theory of Collective Dynamics in Multi-Agent Complex Adaptive Systems</b>	
<i>Neil F. Johnson, Sehyo C. Choe, Sean Gourley, Timothy Jarrett, Pak Ming Hui</i> . . . . .	427
1 Introduction . . . . .	427
2 B-A-R (Binary Agent Resource) Systems . . . . .	428
3 Crowd-Anticrowd Formalism . . . . .	430

## XVIII Contents

4 Implementation of Crowd-Anticrowd Theory .....	434
5 Conclusion and Discussions .....	437
References .....	438

### Modified Mode Coupling Theory of Glassy Dynamics Generated by Entanglement

<i>Rolf Schilling</i> .....	439
1 Introduction .....	439
2 Microscopic Theory of a Discontinuous Glass Transition .....	443
3 Discussion and Conclusions .....	448
References .....	450

---

## Part VIII Spin Systems

---

### Hydrogen and Magnetism in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$

<i>Sebastian T. B. Goennenwein, Thomas A. Wassner, Hans Huebl, Achim Koeder, Wladimir Schoch, Andreas Waag, Jan B. Philipp, Matthias Opel, Rudolf Gross, Martin Stutzmann, Martin S. Brandt</i> .....	453
1 Sample Preparation and Hydrogenation .....	454
2 Hydrogen and Group-II Acceptors in GaAs .....	455
3 Hydrogen in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ : Experimental Results .....	456
3.1 Secondary Ion Mass Spectroscopy .....	456
3.2 Electronic Transport .....	457
3.3 Local Vibrational Mode Spectroscopy .....	458
3.4 Thermal Effusion .....	460
4 Hydrogen and Magnetism in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ .....	460
5 Summary and Outlook .....	463
References .....	464

### Resonant Soft Energy X-Ray Diffraction of Charge, Spin and Orbital Ordering

<i>Peter D. Hatton, Stuart B. Wilkins, Thomas A. W. Beale, Tarnjit Johal, Peter Bencok, D. Prabhakaran, Andrew T. Boothroyd</i> .....	467
1 Introduction .....	467
2 Charge Scattering .....	468
3 Magnetic Diffraction .....	469
4 Magnetic and Orbital Ordering .....	470
5 Conclusions .....	477
References .....	477

### Spin Switching in Mesoscopic Ring Magnets

<i>M. Kläui</i> .....	479
1 Introduction .....	479
2 Switching Type Phase Diagram .....	481

2.1	Thickness Variation . . . . .	481
2.2	Width Variation . . . . .	481
2.3	Outer Diameter Variation . . . . .	482
2.4	Switching Type Phase Diagram for Multi-step Switching . . . . .	483
3	Switching Field Values . . . . .	483
3.1	Outer Diameter Variation . . . . .	483
3.2	Thickness Variation . . . . .	486
3.3	Width Variation . . . . .	487
4	Conclusions . . . . .	489
	References . . . . .	490

## **Modified Magnetic Properties**

### **of Paramagnetic (Zn,Mn)VI at Reduced Dimensions**

<i>P. J. Klar, L. Chen, W. Heimbrod, F. J. Brieler, M. Fröba, T. Kurz, H.-A. Krug von Nidda, A. Loidl . . . . .</i>	491	
1	Template-Directed Growth of Ordered Arrays of (II,Mn)VI Nanostructures . . . . .	492
2	Magnetic Properties . . . . .	493
2.1	EPR Spectra . . . . .	493
2.2	Analysis of the Curie-Weiss Parameter . . . . .	495
2.3	Analysis of the EPR Line Width . . . . .	498
3	Concluding Remarks . . . . .	501
	References . . . . .	502

## **Disorder Effects in Diluted Magnetic Semiconductors**

<i>Carsten Timm, Felix von Oppen . . . . .</i>	503	
1	Introduction . . . . .	503
2	Defect Positions, Growth, and Annealing . . . . .	504
2.1	Defect Clusters . . . . .	505
2.2	Disorder Potential . . . . .	507
3	Hole States . . . . .	508
4	Magnetic Order . . . . .	511
5	Summary . . . . .	513
	References . . . . .	514

## **Magnetic Anisotropy in Ferromagnetic III-Mn-V Semiconductors: Issues and Observations**

<i>J. K. Furdyna, X. Liu, T. Wojtowicz, W. L. Lim, U. Welp, V. K. Vlasko-Vlasov . . . . .</i>	515	
1	Introduction . . . . .	515
2	Uniaxial and Cubic Magnetic Anisotropy . . . . .	516
3	Ferromagnetic Resonance in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ : Effects of Magnetic Anisotropy . . . . .	518
4	Ferromagnetic Domains and the Giant Planar Hall Effect . . . . .	521

5	Manipulation of Magnetic Anisotropy by Controlling Hole Concentration .....	524
6	Closing Remarks .....	528
	References .....	529

---

**Part IX Contact and Interface Quantum Phenomena**

---

**Singlet-Triplet Mixing****in Superconductor-Ferromagnet Hybrid Devices**

<i>Matthias Eschrig, J. Kopu, A. Konstandin, J. C. Cuevas, M. Fogelström, Gerd Schön</i> .....	533
1 Introduction .....	533
2 Basic Equations: Ballistic Case .....	534
2.1 Transport Equations .....	534
2.2 Boundary Conditions .....	535
2.3 Riccati Parameterization.....	538
3 Results: Ballistic Case .....	539
4 Basic Equations: Diffusive Case .....	541
4.1 Transport Equations .....	541
4.2 Boundary Conditions .....	541
4.3 Riccati Parameterization.....	542
5 Results: Diffusive Case .....	542
6 Conclusions .....	544
References .....	544

**Neutron Scattering Investigations of High  $T_c$  Compounds**

<i>Lothar Pintschovius</i> .....	547
1 Introduction .....	547
2 Magnetic Excitations .....	548
3 Lattice Vibrations .....	553
4 Concluding Remarks .....	556
References .....	557
<b>Index</b> .....	559

## Part I

### Solar Cells

# Slow Solar Ascent

Hans J. Queisser

Max-Planck-Institute for Solid State Research  
70506 Stuttgart, Germany  
[queisser@fkf.mpg.de](mailto:queisser@fkf.mpg.de)

**Abstract.** Limited attention arose from the minuscule note by Chapin, Fuller, and Pearson on a new solar cell, as received on January 11, 1954 by the Journal of Applied Physics. Pentagon engineers sneered at this feeble power source; trials failed for rural telephony, but solar cells successfully powered the first Soviet Sputnik. This “Sputnik shock” suddenly rekindled active research in the United States. The initial theory for solar cells relied on mere extensions for the forward current-voltage relations of  $p$ - $n$  junctions. Later work simply attempted reductions of all loss mechanisms. The optimal semiconductor was to be found. A thermodynamic theory was proposed on the basis of detailed balance and a restriction to radiative recombination. Silicon proved to be close to maximal efficiency.

## 1 Introduction: Junctions

Germanium was a most appropriate choice as the material for the initial research toward new amplifying solid state devices. The researchers at the Bell Telephone Laboratories chose this simple, cubic, elemental semiconductor, although their engineering colleagues at the Western Electric Company did not understand this strategy [1]. Why work with this exotic, expensive material, while other semiconductors, such as selenium for rectifiers, or cadmium sulfide for photodetectors, were already in industrial use? Yet, William Shockley’s tenet for studying “the simplest case” prevailed at the labs in Murray Hill, New Jersey. Earlier war time research [2] had already shown that Ge was useful for Radar detector diodes. Germanium was indeed expensive, and the source in the Belgian Congo seemed insecure because of political strife. But Ge was easy to purify; oxygen – for example – presented no real problem.

The  $p$ - $n$  junctions in germanium behaved beautifully. The relation between current  $I$  and applied voltage  $V$  (in a simple abrupt junction) was correctly described [3] by the “ideal rectifier equation”

$$I = I_0 \left( e^{qV/kT} - 1 \right), \quad (1)$$

where  $q$  is the electron charge and  $kT$  the thermal energy. The prefactor  $I_0$  contains only well defined semiconductor parameters, which are accessible via other direct measurements, such as lifetimes  $\tau$ , carrier mobilities  $m$ , or



**Fig. 1.** W. Shockley, inventor of the transistor (Nobel Prize 1956) in his “Apricot Barn” office

diffusivities  $D$ , and diffusion lengths  $L$ , and dopant-induced carrier densities  $n_p$ ,  $p_n$ ,

$$I_0 = qD_p p_{n0} \frac{1}{L_p} + qD_n n_{p0} \frac{1}{L_n}. \quad (2)$$

No adjustable parameters were needed to fit the experimental junction data to this theory. This excellent agreement between theory and experiment finally resolved the disgraceful “*detector problem*” as Walter Schottky had called our embarrassing ignorance concerning rectifiers [4]. With these Ge junctions, semiconductors succeeded to finally establish the solid state to enter physics as a legitimate subject. Reproducibility and interpretation were now firmly secured – extinguished was the awful reputation of solids being hampered by uncontrollable “dirt effects”

## 2 Silicon Conquest

Soon, however, it became apparent that Ge suffered serious problems in applications. The low band gap of Ge, being about 0.7 eV at room temperature, causes the intrinsic carrier density  $n_i$  to be unpleasantly high (at room temperature,  $n_i = 2 \cdot 10^{13} \text{ cm}^{-3}$ ). Slight temperature increases, such as by leaving your radio in the car, parked in sunshine, already flooded the transistors with intrinsic carriers, thus overcoming the intended doping structures. An element with a higher band gap was urgently needed. The clear contender was silicon, with a band gap of about 1.1 eV at room temperature and a comfortable intrinsic density of  $n_i = 1.5 \cdot 10^{10} \text{ cm}^{-3}$ . Silicon, however, presented substantially more stringent requirements for purification; the oxide of silicon, for example,

is extremely stable. Other impurities, especially boron, are truly difficult to remove. Much technology development ensued.

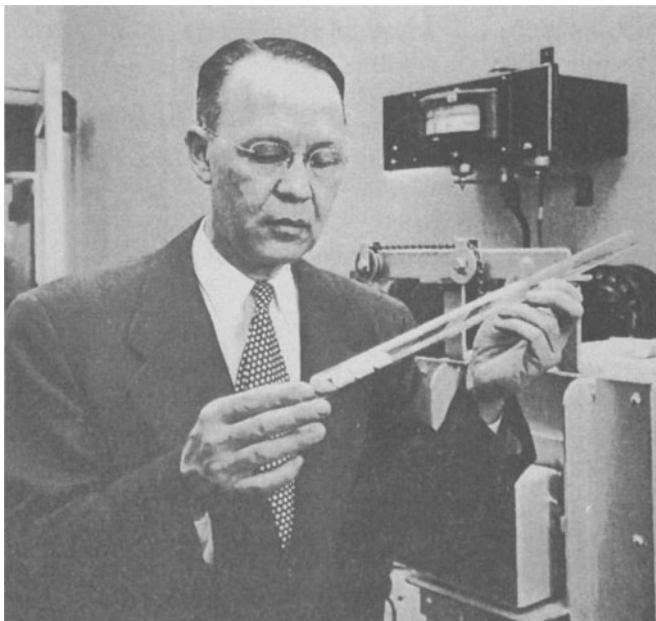
Transistors, both with Ge and Si, were initially fabricated with a simple and inexpensive alloying technique. A thin slab of semiconductor was covered on both sides with dopant pills, which were easily alloyed into the base. A *p-n-p* or an *n-p-n* structure was thus conveniently produced. This simple technology reduced the price of Ge transistors, which was extremely important for semiconductors to invade the territory of the reigning, well-established electron tubes [5]. Alloying, however, even with good control of temperature and time, is not really a well defined method for semiconductor applications. The junctions were irregular; their depths were not at all under close control. Very thin base layers are required for high-frequency operation by facilitating quick carrier transit from emitter to collector, alloying was not sufficient for this requirement. Severe restrictions arose in obtaining transistors for high frequencies.

A novel method was needed for introducing donors and acceptors into the bulk of silicon. Already at that early time, one thought of implanting ions by means of a particle accelerator or by a suitable gas discharge. Implantation seemed nice, because it was a nonthermal, thus low-temperature process. The cost, however, appeared exorbitant. Discharges seemed irreproducible. The only viable solution out of this dilemma was to employ diffusion of the dopants from a gas source inside a well-controlled furnace ambient.

Calvin Fuller and his associates, especially J. A. Ditzenberger, tackled the diffusion technology [6]. They were assisted by systematic studies of the diffusivities of dopants in Si as functions of diffusion temperature. In 1955, M. Tanenbaum and D. E. Thomas succeeded to fabricate a diffused Si transistor with superior behavior at high frequencies [7]. Note that this achievement came almost one decade after the first Ge point contact transistor! It should take many years still for the Si transistors to dominate the competition by Ge and the old vacuum tubes [5]. Gerald Pearson and Calvin Fuller, both at Bell Labs in Murray Hill, reported about a silicon power transistor – and lightning protector – in 1954 [8]. Handling of higher power is another appreciable advantage of the much more stable Si, as compared to Ge.

### 3 A Novel Solar Energy Converter

Calvin Fuller and Gerald Pearson now had shallow diffused junctions in Si wafers at their disposal. Comparatively large areas – for that time, almost  $6\text{ cm}^2$  – were available. They realized that these shallow junctions in such “large” Si wafers might be useful for a “*new solar battery*”. Daryl Chapin was consulted, he was the expert on battery power sources. The three authors wrote a paper around Christmas time of 1953; it was received on January 11, 1954 by the Journal of Applied Physics [9]. A very small note with just one diagram appeared; it is the birthday mark for the semiconductor



**Fig. 2.** C.S. Fuller holding a tube used to prepare silicon for use in the Bell solar battery

junction photovoltaic cell. The efficiency of nearly 5 % for this cell, however crude in today's terms, was nevertheless far superior to all previous radiation converters.

Photovoltaic effects near junctions had earlier caused great excitement at Bell Labs. When Russell Ohl probed a silicon rod with a light spot, he was surprised to measure voltages in the tens of millivolt regime, much larger than could reasonably be expected [10]. The interpretation – after apparently very vivid discussions – was eventually clear: there must be some sort of a junction generating carrier separation. It was then when the terminology *p-type* and *n-type* was coined. Pearson and Fuller, of course, knew of this effect of photons on a semiconductor junction.

American Telephone and Telegraph, the mother corporation of Bell Labs and Western Electric, was – at that time – a government-controlled monopoly. The Bell Laboratories were involved in military developments, for example in nuclear technology at the Sandia Labs in Albuquerque, New Mexico or in tracking missiles from the Kwajalein atoll in the middle of the Pacific Ocean. Any new invention had to be presented to the military authorities. Hence, a delegation went to the Pentagon to demonstrate the new solar cell [11]. The engineers in uniform, however, were not at all interested in this feeble power source. Magnanimously, they granted permission to openly publish and to apply for patents.

The Bell Telephone and Telegraph System tried to utilize the new device. The Southern Bell affiliate performed field test in the vicinity of the little town of Americus, Georgia – a sun-rich region. Using this renewable energy source seemed a reasonable proposition: one might save the cost of replacing batteries on the rural telephone poles. The tests did not yield positive results. One of the troubles were the opaque white excrements left by the birds on these apparently attractive black surfaces. Nevertheless, the city of Americus still proudly claims on its web site to be the early user of solar energy. The inventor trio of the cell, Chapin, Fuller, and Pearson, no longer worked on this device. Other things dominated, especially understanding and using diffusion phenomena for faster and more powerful silicon transistors as well as for other semiconductor materials [12]. Gerald Pearson wrote an article in the physics-teacher's journal [13] and presented not exactly a highly promising picture; competitive cost would be difficult to achieve for photovoltaics against conventional fuels. The ascent of the new device was certainly extremely slow.

## 4 Signals from Space

One group, however, secretly kept unbridled, keen interest in solar cells. The physicists and engineers in the Soviet Union were preparing their first little satellite to be launched into orbit. A power source was required. Cost and efficiency were of minor concern; the light weight was all decisive. Later, it became known that cells were smuggled into the USSR, and soon their own excellent cells were fabricated, usually with *n-type* diffusions onto *p-type* base silicon [14]. The beeping *Sputnik* demonstrated also an audible, brilliant success of solar energy conversion by semiconductors.

This unexpected strategic importance of the diffused silicon solar cell by the cold war foe created hectic efforts in the United States. This “*Sputnik shock*” of 1957/58 also rekindled feverish interest in solar cells. Suppliers were encouraged to provide cells. The military laboratories undertook their own research and development and rather generously offered monetary incentives to anybody willing to perform science and engineering on photovoltaics. The United States Army concentrated their efforts at their Signal Corps Laboratory in Fort Monmouth, New Jersey, while the US Air Force designated especially their laboratories at the Wright-Patterson Base in Dayton, Ohio. Funds flowed abundantly. The most important questions touched upon the choice for the optimal semiconducting material and the challenge of achieving high efficiencies. There were competitors, however: small nuclear reactors were seriously proposed as alternate energy sources, but they soon were forced to exit the race.

#### 4.1 Early Interpretation

Understanding the function of a solar cell proved quite easy. Morton B. Prince of the Bell Laboratories proposed an early answer [15]. Sunlight is absorbed and provides excess carrier pairs. Electrons diffuse to their region of minimal energy, which is the *n-type* portion; holes tend towards the *p-type* portion of the junction. This supply of carriers beyond thermal equilibrium is equivalent to a negative, “reverse” current. Hence, the current-voltage relation under illumination is simply

$$I_{\text{total}} = I_{\text{phot}} + I(V), \quad (3)$$

where the negative  $I_{\text{phot}}$  is subtracted from the junction current  $I(V)$ , as described by Eq. (1). The characteristic is merely shifted downward on an  $I(V)$  plot. A portion of the curve linking  $I$  and  $V$  will now be in the fourth quadrant (positive  $V$ , negative  $I$ -values). This position indicates that the device is able to supply electric power to an external load. A rectangle may be constructed to give maximal output. Two new important quantities arise: First, an open-circuit voltage  $V_{\text{op}}$ , which is  $V(I = 0)$  and, secondly, a short-circuit current  $I_{\text{sh}}$ , being  $I(V = 0)$ . Morton Prince provided some simple estimates of achievable conversion efficiency.

#### 4.2 Nonideal Junctions

While low-gap germanium so neatly obeys the ideal rectifier equation of Eq. (1), almost all the other semiconductors behave erratically, certainly they display nonideal junction behavior. A much higher current – for given voltage  $V$  – is flowing across the junction, and only at fairly high voltages does the current approach the predicted values from Eq. (1). There are current pathways across the junction in addition to the thermally activated carrier flow over the junction’s potential barrier. An entirely empirical correction to the rectifier equation is usually done by means of a “junction quality-parameter”  $n_j$ , which changes the exponential from  $\exp(qV/kT)$  to  $\exp(qV/n_j kT)$ . Notice the misnomer: an ideal junction is indicated by  $n_j = 1$ ; nonideal junctions need a force-fit with  $n_j(V) > 1$ .

This nonideality implies serious shortcomings for solar cell junctions, since  $n_j > 1$  means a shift of the current/voltage relation towards smaller voltages for a given current. The open circuit voltage  $V_{\text{op}}$  is reduced, consequently the energy conversion efficiency  $\eta$  falls. These effects complicated estimates for the efficiency. The researchers at the Princeton RCA Laboratories, headed by Paul Rappaport, considered semiconductors of the II-VI or III-V groups as particularly efficient, but these materials with their bandgaps typically above 1.5 eV featured strong nonidealities [16], which were difficult to ascertain.

Engineers working with solar cells and trying to improve efficiencies began to rely on very pragmatic estimates. Losses had simply to be reduced. Thus, all possible loss mechanisms were listed and recipes were sought to minimize

them [17]. A definitive recognition of the optimal material and the value for the maximal efficiency were still missing in the late Fifties of the last century.

### 4.3 Silicon Valley

William Shockley said that he “had seen his name often enough in the Physical Review, now wanted it in the Wall Street Journal” (11, 18). He moved to Mountain View in California’s Santa Clara County, rented an old apricot barn on 391 South San Antonio Road and started a laboratory. This site is now established as the “cradle of Silicon Valley”; Jacques Beaudouin and myself got the City Council of Mountain View to put a bronze plaque there. Shockley hired excellent coworkers, such as Robert Noyce or Gordon Moore, who soon deserted him in a dramatic secession to form Fairchild Semiconductors, later the Intel Corporation. These “Treacherous Eight” were convinced that silicon transistors might be produced profitably, while Shockley insisted that his company ought to make switching diodes [18]. He favored a four-layer  $p-n-p-n$ -“Shockley diode”, which would only work with silicon, since here the nonideality was a major prerequisite [19],

The forward  $I(V)$  characteristic of silicon was therefore essential for Shockley’s plans. He, Noyce, and Sah published a major paper, in which they showed that the nonideality of Si was caused by defects within the space charge layer of the junction [20]. Carriers need not surmount the junction barrier, they could reach the defects, recombine there and thus provide extra currents at low voltages. The “quality parameter” could reach  $n_j = 2$  in the limit.

When I joined Shockley’s laboratory in 1959, solar cell money was aplenty, and silicon already was the preferred material [18]. We had two government contracts to study cells. Shockley had really no intention to fabricate any solar cells, but he urgently wanted to research junction behavior and elements of silicon technology. My first project – financed from Fort Monmouth – was to make a multijunction solar cell out of very thin (60  $\mu\text{m}$ ) Si slabs, using deep Ga diffusions to establish a series-connected cell, yielding voltages close to ten volts [21]. This contraption was really the first “integrated silicon device”, but done with a technology of little general value. The second project was much more scientific: the Air Force at Dayton wanted a reliable statement about optimal material and maximal efficiency.

### 4.4 Detailed Balance

We on San Antonio Road did not like the naïve treatment of merely minimizing energy conversion losses. This desperate strategy to us resembled the attempts of reaching better steam engines before Carnot presented a true thermodynamic law for maximal efficiency. A similar thermodynamic theory was needed for solar cells. The principle of detailed balance was invoked, stating that a cell at room temperature must also act like a black-body emitter. The

ideal case would be reached if the semiconductor would only permit radiative carrier recombination; all non-radiative processes would be suppressed. This approach followed along the lines of estimating carrier lifetimes, done by van Roosbroeck and Shockley [22]; the observed long lifetime of excess minority carriers was a pleasant surprise for the early transistor makers and had to be understood in its limits. I had no computer for my many calculations, did everything by hand and by means of a tricky black-body slide rule [23].

Our paper on detailed balance and the theoretical maximum energy efficiency of about 30% was submitted to the Journal of Applied Physics. The stern reviewers rejected it as being just “a rehash”. I remember how irate my boss got, but we sat down, enhanced and rewrote the paper, which was finally accepted [23]. Few people quoted us, however. Again, the ascent was slow. Yet, our estimate that silicon was well within the rather broad maximum of the curve efficiency  $\eta$  versus band gap  $E_g$  was useful; the previous favorites from the compound semiconductors lost their status. Our prediction that materials with predominantly radiative recombination meant that the direct-gap GaAs ought to be a supreme candidate, which was not yet observed initially [24], but became a shining reality as soon as the nonradiative surface recombination processes in GaAs were suppressed [25].

I shall never forget the icy reception I received when traveling to our sponsor and obediently submitting our final report. The people at Wright-Patterson Air Force Base were not amused. The significance of our theoretical approach was not at all appreciated. Worse, I was late in delivering the report, because we had to wait for a chapter on Auger recombination by the expert Peter Landsberg from Britain. Shockley and I had realized that this type of recombination was an unavoidable nonradiative mechanism for high doping levels and therefore we consulted the undisputed specialist [23].

Eventually, our paper did become interesting to the community. It was reprinted several times in collections of solar cell papers [26]. Nowadays, I am understandably pleased to find that our approach is being accepted and very frequently quoted as the standard, correct treatment for solar cell efficiency. My very last conversation with Shockley (in the men’s room of the Stanford Faculty Club) just before his death indeed rang with a pleased satisfaction about our work.

## 5 Third Generation

Contemporary efforts of improving solar cells to lead them to competitiveness are defined by “third generation” cells, which are supposed “to exceed the Shockley-Queisser limit” [27]. Various ways are being examined. Concentration of the solar input by lenses and multijunction cells are being considered and will be discussed elsewhere in this volume. Addition of defect levels within the gap are still estimated to be useful for capturing low-energy photons, although a very early careful treatment by Güttler and myself [28] negated

such proposals. Juergen Werner and colleagues at the Stuttgart Max-Planck-Institute considered multipair carrier generation to utilize the excess energy of blue solar photons; but the expectable improvement seems small [29]. Future research and development on solar energy converters must be aimed towards further cost reductions.

### Acknowledgements

I wish to extend my gratitude to the late William Shockley for his strict yet benevolent training and cooperation. I am thankful for many friendly encounters with smiling, cigar-smoking Gerald Pearson, especially at Stanford, and with the righteous Calvin Fuller, with whom I really enjoyed collaborations at Murray Hill. The lively discussions with Paul Rappaport and with my contract monitors were wonderfully educational. Later, at Frankfurt University, it was a delight to work with the imaginative Gerhard T. Gütler. For our photovoltaics work at Stuttgart, I thank the undaunted Juergen Werner, Sabine Kolodinski, Uwe Rau, Ralf Bergmann, and Rolf Brendel and many younger colleagues for their exuberant enthusiasm and their successful efforts.

### References

1. W. Shockley, IEE Trans. El. Dev., **ED-23**, 7 (1976).
2. L. N. Ridenour, *Crystal Rectifiers*, (MIT Radiation Lab Series, McGraw-Hill, NY, 1948).
3. W. Shockley, *Electrons and Holes in Semiconductor*, (van Nostrand, Princeton, 1953).
4. H. J. Queisser, Siemens Forschg.u. Entwickl. Berichte. **15**, 272 (1987).
5. E. Braun and S. Macdonald, *Revolution in Miniature*, (Cambridge, 1978).
6. C. S. Fuller and J. A. Ditzenberger, J. Appl. Phys. **25**, 760 (1954).
7. M. Tanenbaum and D. E. Thomas, Bell Syst. Tech. J. **35**, 1 (1956).
8. G. L. Pearson and C. S. Fuller, Proc. IRE **72**, 760 (1954).
9. D. M. Chapin, C. S. Fuller, and G. L. Pearson, J. Appl. Phys. **25**, 676 (1954).
10. see *History of Eng. and Sci. in the Bell System*, S. Millman (ed.) ATT Bell Labs. (1983).
11. personal communication by W. Shockley to H. J. Queisser,
12. H. J. Queisser and C. S. Fuller, J. Appl. Phys. **37**, 4895 (1966).
13. G. L. Pearson, Am. J. Physics **25**, 591 (1957).
14. private communications by V. Vavilov to H. J. Queisser.
15. M. B. Prince, J. Appl. Phys. **26**, 534 (1955).
16. J. J. Wysocki and P. Rappaport, J. Appl. Phys. **31**, 571 (1960).
17. M. Wolf, Proc. IRE **48**, 1246 (1960).
18. H. J. Queisser, *Kristallene Krisen* (Piper, München, 1985); engl. version: *The Conquest of the Microchip*, (Harvard Press, Cambridge, 1985).
19. K. Hubner in *Semiconductor Silicon*, H. R. Huff, (ed.) Proc. Electrochem. Soc. **98-1** (Pennington, New Jersey, 1998), p. 49.
20. C. T. Sah, R. N. Noyce, and W. Shockley, Proc. IRE **45**, 1228 (1957).

21. an illustration and the reference are in: H. J. Queisser, Physica E **14**, 1 (2002).
22. W. van Roosbroeck and W. Shockley, Phys. Rev. **94**, 1558 (1954).
23. W. Shockley and H. J. Queisser, J. Appl. Phys. **32**, 510 (1960).
24. Discussion at APS March Meeting, Cleveland (Ohio), Bull. Am. Phys. Soc. **5**, 160 (1960).
25. J. M. Woodall and H. J. Hovel, Appl. Phys. Lett. **30**, 492 (1977).
26. reprinted *e.g.* in: S. M. Sze, *Semiconductor Devices: Pioneering Papers* (World Scientific, Singapore, 1991).
27. J. H. Werner discusses the "Third Generation", Adv. Sol. St. Phys., this volume (2004).
28. G. Güttsler and H. J. Queisser, Energy Conv. **10**, 51 (1970); se also J. Appl. Phys. 4994 (1969).
29. J. H. Werner, S. Kolodinski, and H. J. Queisser, Phys. Rev. Lett. **72**, 3851 (1994).

# Interface Analysis of Thin Film Solar Cells Using Photoelectron Spectroscopy

Andreas Klein

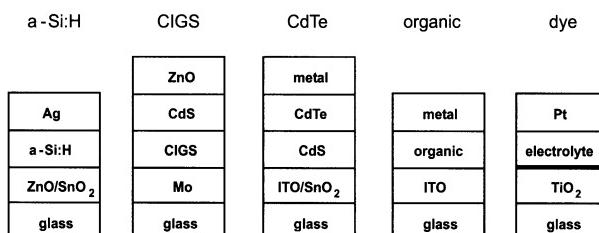
Darmstadt University of Technology, Institute of Materials Science  
Surface Science Division, Petersenstrasse 23, 64287 Darmstadt, Germany

**Abstract.** Interfaces are important for the efficiencies of thin film solar cells. In particular for polycrystalline chalcogenide semiconductors as Cu(In,Ga)(S,Se)<sub>2</sub> and CdTe the existing physical concepts, which describe the electronic properties of semiconductor interfaces, are not sufficient. The increased complexity is mostly due to the non-abruptness of the interfaces and the strong tendency for the formation of defects. Photoelectron spectroscopy can significantly contribute to the understanding of the mechanisms governing the properties of semiconductor interfaces in thin film solar cells. The experimental approach using integrated surface analysis and thin film deposition systems and selected results will be presented.

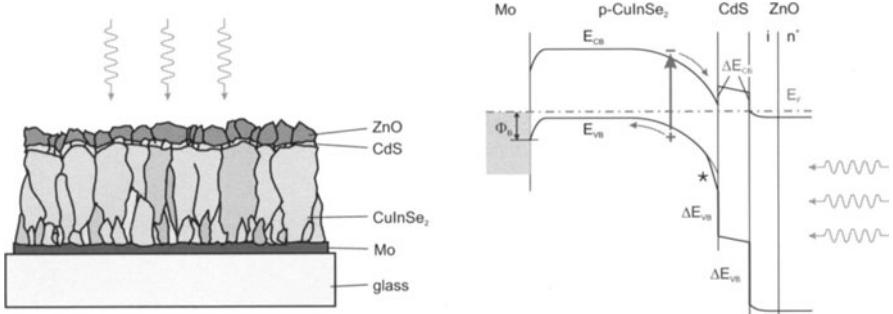
## 1 Interfaces in Thin Film Solar Cells

Various types of thin film solar cells with noticeable efficiencies are known. The layer sequences of the amorphous silicon (a-Si), Cu(In,Ga)(S,Se)<sub>2</sub> (CIGS), CdTe, organic and dye sensitized solar cells (dye) are schematically shown in Fig. 1. Important interfaces are: semiconductor / metal (Schottky contact), semiconductor / semiconductor (heterocontact) and semiconductor / TCO (transparent conducting oxide).

Many of the interfaces shown in Fig. 1 have been investigated using photoelectron spectroscopy giving first ideas on the important parameters. Since the efficiency of all thin film solar cells depends strongly on the techniques used for film deposition, it has to be expected that also the interface properties are changing. For example, the CIGS thin film solar cell shows highest efficiencies with a co-evaporated Cu(In,Ga)(S,Se)<sub>2</sub> light absorber, a chemical bath deposited CdS buffer layer and a sputtered bilayer of undoped and aluminum doped ZnO (window layer) [1].



**Fig. 1.** Layer sequences for different thin film solar cells



**Fig. 2.** Cross section of a Cu(In,Ga)(S,Se)<sub>2</sub> thin film solar cell: schematic setup (left) and energy band diagram according to Ref. [2] (right). The CIGS layer has an increased band gap close to the interface with CdS as a result of a Cu-deficient surface layer (\*) [3]. ( $E_F$ : Fermi level,  $E_{VB}$ : valence band maximum,  $E_{CB}$ : conduction band minimum,  $\Delta E_{VB}$ : valence band offset,  $\Delta E_{CB}$ : conduction band offset,  $\Phi_B$  barrier height).

Important properties of interfaces are the barrier heights (band discontinuities) and the density and energetic distribution of defects. Together with the doping of the semiconductors, these determine the potential distribution throughout the cell. To illustrate this, a schematic setup and the energy band diagram of a CIGS thin film solar cell is shown in Fig. 2.

The solar cell is made of a number of layers of different materials and essentially contains three interfaces:

- The contact Mo / Cu(In,Ga)(S,Se)<sub>2</sub>

This serves as electric contact to the p-doped Cu(In,Ga)(S,Se)<sub>2</sub> layer. It is typically assumed that the barrier height  $\Phi_B$  should be as low possible for good ohmic contact (see also section 5.1)

- The contact Cu(In,Ga)(S,Se)<sub>2</sub> / CdS

This interface is most important for high conversion efficiency since it is in the area of maximum light absorption. Recombination has to be minimized, which is best achieved by a Fermi level close to the CIGS conduction band at the interface and a small positive conduction band offset as indicated in Fig. 2

- The contact CdS / ZnO

ZnO is a transparent semiconductor, which can be degenerately doped n-type with Al. High conversion efficiencies are only obtained when a nominally undoped ZnO is deposited onto CdS

## 2 Band Alignment

Considerable experimental and theoretical work has been dedicated in the past 30 years to understand semiconductor interface formation and particu-

larly the microscopic origin of the barrier heights for charge transport. Extensive descriptions can be found in [4,5,6,7]. Since the band discontinuities are important for the electronic properties of the interfaces, it was also desired to predict them for a certain material combination. The first idea was to align the vacuum level  $E_{vac}$ , which results in a conduction band offset  $\Delta E_{CB}$  equal to the difference in the electron affinities ( $E_{vac} - E_{CB}$ ) of the two materials and in a valence band offset  $\Delta E_{VB}$  equal to the difference in ionization potential ( $E_{vac} - E_{VB}$ ). This model is known as electron affinity rule (EAR). There is no principle difference between semiconductor heterocontacts and Schottky contacts. For metals, electron affinity, ionization potential and work function ( $E_{vac} - E_F$ ) are the same and the conduction / valence band offset correspond to the Schottky barrier height for electrons / holes (Mott-Schottky model).

The failure of the electron affinity rule to predict band alignment at semiconductor interfaces has been recognized very early. The Schottky barrier heights for most important semiconductors depend only slightly on the contact metal, which is known as *Fermi level pinning*. Its origin is related to defect states in the semiconductor close to the interface. Together with the charge at the metal surface, the charge in these *interface states* lead to an electric dipole and a corresponding dipole potential, which shifts the band edges of the two materials relative to each other.

The microscopic origin of the interfaces states has been related to metal-induced-gap-states (MIGS) (see e.g. [4,8,9]). Even for atomically abrupt interfaces, the wave functions of the metal penetrate into the semiconductor band gap. The degree of Fermi level pinning is determined by the density of the induced states, which depends on the ionicity of the semiconductor [9]. Covalently bonded semiconductors as silicon and GaAs generally show a stronger Fermi level pinning than more polar bonded semiconductors, as e.g. ZnSe and oxides. For highly pinned semiconductors, the barrier height does not depend on the metal and is given by the position of the *charge neutrality level* of the semiconductor. This can be theoretically calculated using different approaches as branch point energies [8] or dielectric midgap energies [10]. Again there is no fundamental difference between semiconductor / metal and semiconductor heterocontacts. For the latter, the band alignment is given by alignment of the charge neutrality levels.

Unfortunately the charge neutrality levels have not been calculated for a large number of materials in thin film solar cells (see Fig. 1). In addition, the II-VI compound semiconductors CdTe, CdS and its multinary analogue Cu(In,Ga)(S,Se)<sub>2</sub> have considerable ionicity and the Fermi level might therefore be not strongly pinned. It is also not clear whether this concept may apply for interfaces with organic materials or electrolytes.

An alternative approach for the explanation of Schottky barrier heights and Fermi level pinning is the *unified defect model* [11]. It associates the defects in the band gap and the corresponding charge neutrality pinning

level to intrinsic defect states of the material, as e.g. antisite defects. The defects are assumed to be created by the heat of adsorption of the metal atoms. A variation of the defect model is the amphoteric defect model by Walukiewicz, which, in conjunction to the self-compensation mechanism in semiconductors, assumes that the defects are created by the movement of the Fermi level during barrier formation [12].

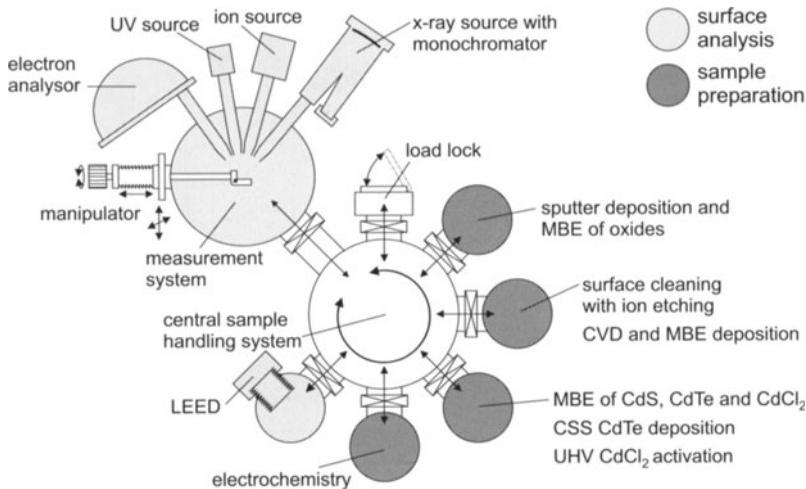
### 3 Photoelectron Spectroscopy

Photoelectron spectroscopy (PES) is a widely used technique for surface and interface analysis [13,14]. The high surface sensitivity is due to the inelastic mean free path of the photoelectrons  $\lambda_e$ , which depends on the electron kinetic energy  $E_{kin}$ . Typical excitation energies for photoelectron spectroscopy are within  $10 - 1500$  eV giving corresponding kinetic energies of the photoelectrons, which then have mean free paths of  $5 - 30$  Å. Most frequently used (standard) applications of PES are elemental and chemical surface analysis [14]. A less well-known feature of PES is its capability to detect electronic surface properties, which traces back to the measurement of binding energies with respect to the Fermi level. Since samples are electrically connected to the spectrometer system, the Fermi level is at a constant energy, which can be determined by a calibration measurement using a metallic sample. For semiconductors, the Fermi level can change within the band gap, leading to changes of all binding energies with Fermi level changes. If for a given sample the binding energy of a core-level is known with respect to the valence band maximum ( $E_{CL} - E_{VB}$ ), the core-level binding energy itself can be directly used as a measure for the position of the Fermi level in the band gap ( $E_F - E_{VB}$ ).

Determination of Schottky barrier heights  $\Phi_B$ , or valence band discontinuities  $\Delta E_{VB}$ , can be performed by following the evolution of the position of the valence band maxima with respect to the Fermi level of substrate and overlayer with increasing thickness of the overlayer [15]. For layer-by-layer growth the attenuation of the substrate intensities is given by the inelastic mean free path of the photoelectrons. Hence, after a film thickness of  $\sim 5$  nm, the substrate emissions are completely extinguished. The study of interface formation with PES thus requires the control of film thickness in the sub-nanometer range.

### 4 Experimental Procedure

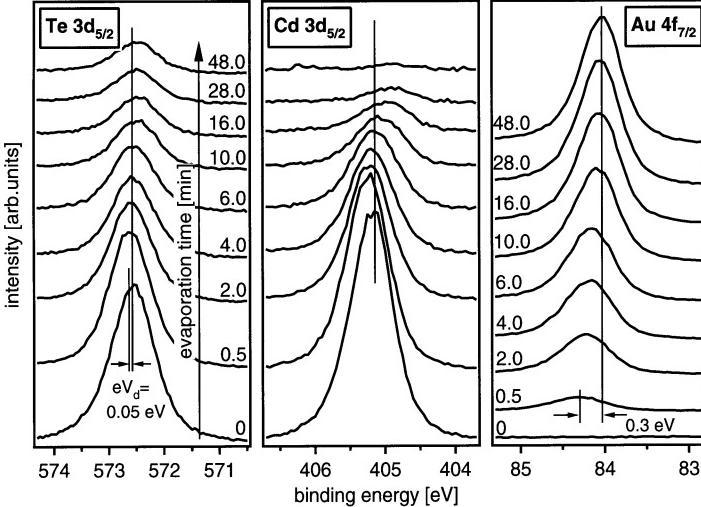
The rapid contamination of surfaces in air requires that the sample transfer between surface preparation or film deposition to the photoelectron spectrometer is performed in vacuum. Since a large variety of interfaces occur in thin film solar cells, different deposition techniques have to be implemented. Consequently the analysis system has to *integrate* several preparation chambers



**Fig. 3.** Layout of the integrated surface analysis and preparation system DAISY-MAT (DArmstadt Integrated SYstem for MATerials Research). A photoelectron spectrometer is connected by a sample handling system to various deposition and surface treatment chambers. Preparation and analysis can be repeatedly performed under controlled ultrahigh vacuum conditions

and a surface analysis tool. The layout of such a system used at Darmstadt University is shown in Fig. 3.

The first step for a well-defined interface experiment is the preparation of a clean substrate surface, which can be achieved by thin film deposition of the material itself. Heating is often not sufficient and sputtering destroys the material surface in most cases. However, preparation of thin films for solar cells can require a long-term know-how and particularly designed deposition systems. Neither the know-how, nor the space for such deposition systems is typically available in interface analysis groups. An elegant way to combine state-of-the-art thin film deposition and interface analysis is the protection of surfaces by means of a material, which can be heated-off from the surface in the vacuum system. We have implemented such a *capping* procedure for Cu(In,Ga)Se<sub>2</sub> together with different groups [16,17,18]. The steps to prepare an uncontaminated surface are (i) deposition of the CIGS layer at high temperature (ii) cooling of the sample in the deposition chamber and deposition of a selenium cap layer (iii) transport and storage of the samples (iv) insertion into the surface analysis system and heating-off (de-capping) of the Se layer at 300°C. The de-capping process does not deteriorate the absorber properties for its use in solar cells [18].



**Fig. 4.** Core-level photoelectron spectra obtained during growth of Au on a polycrystalline CdTe surface

## 5 Example Results

### 5.1 CdTe / Metal Interfaces

The back contact is a severe problem for CdTe thin film solar cells. There is always a barrier at the interface of the (p-type) CdTe absorber with the contact, which shows up in the current-voltage characteristic as the so-called roll-over phenomenon [19]. We have carried out systematic studies of the contact formation of metals with surfaces of polycrystalline CdTe films [20]. An example set of spectra is shown in Fig. 4. The spectra at the bottom are from the uncovered CdTe substrate. With increasing thickness of the Au film, the substrate lines are attenuated, while the Au 4f intensity increases. At very low coverage, there is a small shift of 0.05 eV of both Cd and Te 3d levels to lower binding energies (BE) as a result of a band bending induced by Au deposition. The binding energies of the CdTe are translated to the Fermi level position  $E_F - E_{VB}$  by  $BE(Cd\ 3d) - BE_{CdTe}(VBM) = 404.39\text{ eV}$  and  $BE(Te\ 3d) - BE_{CdTe}(VBM) = 571.75\text{ eV}$ .

For larger Au coverage, the Cd 3d level shifts further to lower binding energies, while the binding energy of the Te 3d level remains constant. The different BE shifts are the result of a chemical decomposition of the substrate, which is also evident from the different attenuation of the Cd and Te core-levels. Even for high Au deposition times, the Te 3d signal is not completely extinguished. This can be explained by a Te layer, which floats on the surface of the growing film. The decomposition of the CdTe substrate is thermodynamically driven by the solution of Cd in the Au film, rather than

**Table 1.** Chemical and electronic properties of different CdTe / metal contacts. Metal deposition has been performed at room temperature conditions;  $\phi_m$ : metal workfunction,  $\Phi_B$ : barrier height; Abbreviations for reactivity are A: no reaction, B: decomposition of the substrate, Cd:M alloy formation, Te segregation at the surface, and C: decomposition of the substrate, Cd<sup>0</sup> disposal at the interface due to Telluride formation

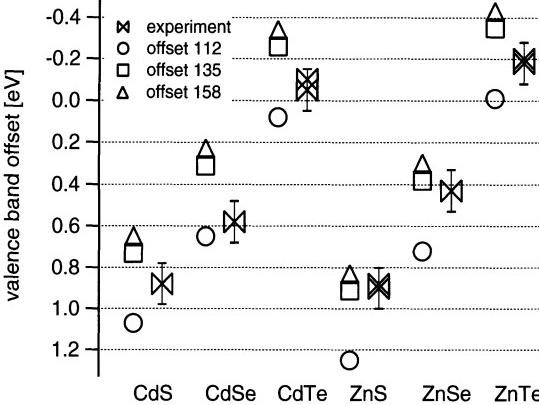
metal	$\phi_m$ [eV]	$\Phi_B$ [eV]	reactivity
Cd	4.2	1.18	A
Au	5.1	0.85	B
Cu	4.65	0.95	B
Pt	5.6	1.0	B
Ti	4.3	1.15	C
V	4.3	1.07	C
W	4.55	1.05	C
Sb	4.55	0.97	C

by the formation of a Au-Te compound. Such chemical reactivities have been identified already in the 1980's [21].

Also other types of reactivity have been observed at CdTe / metal interfaces [20]. Transition metals (TM) strongly tend to form tellurides, hence also leading to a decomposition of the CdTe substrate. In this case sometimes also TM-Cd alloying occurs. A summary of the determined barrier heights and interface reactivities is given in Table 1.

Both barrier heights and reactivities are similar to those reported earlier for CdTe single crystals [22,23]. As reported by Shaw et al. [22] the barrier heights coincide with energy levels of bulk CdTe defects, which are created at the interface in a high concentration by the chemical reactions. Our experiments show that almost all metals lead to the decomposition of the CdTe, releasing elemental Cd. The energy level of interstitial Cd in CdTe has been calculated by Wei and Zhang as  $\sim 1$  eV above the valence band maximum [24]. The agreement between chemical reactivity, Cd<sub>i</sub> defect level and the barrier heights suggests that the barrier height is determined by the intrinsic Cd defect levels as in the unified defect model [11]. Since interface reactivity can not be avoided, it seems hopeless to prepare ohmic back contacts to CdTe by metal deposition. Nevertheless, CdTe thin film solar cells with CdTe / metal back contacts show reasonable efficiencies, indicating good charge transport across the interface despite large barrier heights.

A possible interlayer at the back contact is ZnTe. Its advantage is that higher p-type doping levels can be obtained compared to CdTe [25]. The interface properties of ZnTe are similar to those of CdTe, although a somewhat larger variation of the barrier height ( $\Phi_B = 0.6 - 1$  eV) is observed [23,26]. Electrical measurements show a lower contact resistance of Mo compared to



**Fig. 5.** Experimental valence band offsets between  $\text{CuIn}_{0.7}\text{Ga}_{0.3}\text{Se}_2$  and different II-VI compounds. The CIGS surfaces have been prepared by the de-capping procedure and exhibit a Cu-poor surface composition. The theoretical band offsets ( $hkl$ ) are given for  $\text{Cu}_h(\text{In},\text{Ga})_k\text{Se}_l$  surface composition [28]

Au, although the barrier for CdTe / Mo (1 eV) is significantly larger than the one for CdTe / Au (0.65 eV) [26]. This apparent contradiction between PES and electrical measurements gives clear evidence that current transport at the interface does not occur over the barrier as in the thermionic emission model, but is rather accomplished via the defects introduced by the interface reaction.

## 5.2 CIGS / II-VI Compound Interfaces

The CIGS films used for the de-capping process (see section 4) have been deposited using a slight Cu poor composition. It is known that the surfaces of such films exhibit a Cu-poor  $\text{Cu}(\text{In},\text{Ga})_3\text{Se}_5$  composition [3], which is also found by our measurements after de-capping. The films have been used to systematically determine the band alignment to II-VI compounds, which are used as buffer layers in CIGS solar cells. More details of the experiments can be found elsewhere [17,18,27]. The valence band offsets determined from these experiments, which are highly reproducible, are shown in Fig. 5.

The experimentally determined offsets are compared to ab-initio density functional theory calculations for  $\text{Cu}(\text{In},\text{Ga})\text{Se}_2$  / CdS (112),  $\text{Cu}(\text{In},\text{Ga})_3\text{Se}_5$  / CdS (135) and  $\text{Cu}(\text{In},\text{Ga})_5\text{Se}_8$  / CdS (158) [28]. The lower concentration of Cu atoms in the Cu-deficient compounds leads to a reduced *p-d repulsion*, leading to a lowered valence band maximum position [28]. Hence a different valence band offset has to be expected depending on the surface stoichiometry. From Fig. 5 it is obvious, that the experimental values are close to the theoretical ones and thus give an identical general trend. However, the experimental band offsets can not be directly correlated to the theoretical values for a single composition.

Not all of the theoretical band alignments are directly calculated from a supercell containing the two materials but are rather added band offsets calculated for individual interfaces assuming transitivity of band offsets ( $\Delta E_{VB}[AC] = \Delta E_{VB}[AB] + \Delta E_{VB}[BC]$ ). In addition the experimental error is typically given as  $\pm 0.1$  eV. The observed deviations between experiment and theory might therefore be attributed just to the uncertainty of both. However, because of the high reproducibility of experimental values, which we attribute to the well-defined and reproducible preparation conditions, we estimate a considerably lower experimental error. There are also other observations, which indicate that the deviation between experiment and theory in Fig. 5 have their origin in a modification of the interface structure:

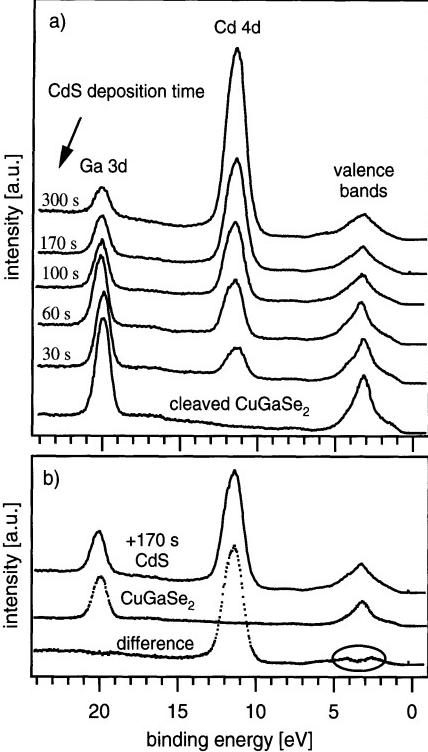
- The valence band offsets between stoichiometric (112) Cu(In,Ga)Se<sub>2</sub> crystals and CdS are very close to those for thin film CIGS with (135) surface composition [17,27,29,30], although theory predicts a 0.3 – 0.4 eV difference [28]. This might be explained by changes in the Cu concentration near the interface. Unfortunately, no systematic comparisons between Cu-rich and Cu-poor surface conditions for similarly prepared substrates are available yet. Work is in progress to clarify this point.
- A detailed comparison of the band offsets at CuInSe<sub>2</sub> / CdS and CuGaSe<sub>2</sub> / CdS indicates an opposite trend compared to theory [17]. This might also be explained by a chemical modification of the interface of similar origin, although a detailed explanation can not be given yet.

The band alignments calculated by Wei et al. generally give very good agreement to a number of experiments. However, the calculations are so far limited to atomically abrupt interfaces between two well-defined crystallographic phases. The calculated band alignments do not take into account the actual chemical structure of the interface. In particular if changes of the Cu concentration occur at the CIGS surface close to the CIGS / II-VI interface, this might strongly affect the valence band offset.

There is direct evidence from photoemission measurements for a removal of Cu from CIGS surfaces with 112 stoichiometry, which is shown in Fig. 6 [29,32,33]. Deposition of CdS or Na onto these surfaces leads to an upward Fermi level shift, which is limited to  $E_F - E_{VB} \approx 0.85$  eV. The observation of comparable Fermi level positions in conjunction with a reduction of the Cu concentration from the surface has led to the conclusion, that the effect has an electronic origin and is related to a self-compensation mechanism of the CIGS material [29,32].

### 5.3 TCO Interfaces

The interfaces of transparent conducting oxides are also important for all kind of thin film solar cells. The TCOs are typically used as degenerately doped n-type semiconductors. Nevertheless, in amorphous silicon or in organic solar



**Fig. 6.** Photoelectron spectra taken during CdS deposition onto a CuGaSe<sub>2</sub> single crystal fractured in vacuum [29]. Spectra were taken using  $h\nu = 80$  eV synchrotron radiation. The bottom part shows the spectra of the clean sample, the spectra taken after a CdS deposition time of 170 s and the corresponding difference spectra. The difference spectra show a dip at a binding energy of  $\sim 3$  eV, which corresponds to the maximum of the Cu 3d partial density of states [31]. The dip is therefore attributed to a diffusion of Cu from the surface into the bulk of the CuGaSe<sub>2</sub> [29]. Since similar effects have also been observed for other CIGS interfaces, the origin of Cu diffusion has been attributed to the shift of the Fermi level closer to the conduction band minimum, driven by interface formation [32]

cells, they are also used to contact the p-type semiconductor. This already shows that the TCO interfaces are also more complex than expected for degenerately doped semiconductors, which, on the first sight, might be treated like a metal. Experiments on reactively evaporated In<sub>2</sub>O<sub>3</sub> films clearly show that the surface Fermi level lies  $\sim 1$  eV below the conduction band minimum and not above, as expected for degenerate doping [34]. This has been very recently also found for ITO films [35]. While SnO<sub>2</sub> shows a comparable behavior [36], a Fermi level position above the conduction band minimum can be found at surfaces of Al doped ZnO [37].

The Fermi level at the TCO surfaces has considerable impact on the function of the contact. At the SnO<sub>2</sub> / CdS interface e.g., it leads to low Fermi level position in the CdS, which can seriously reduce the photovoltage [36]. One of the modifications induced by the CdCl<sub>2</sub> activation, which is essential for high conversion efficiencies, is an upward shift of the Fermi level in the CdS [38].

## 6 Conclusions

Photoelectron spectroscopy is a highly useful tool to study semiconductor interface formation. Its capability to simultaneously detect chemical and elec-

tronic properties can be exploited for understanding the fundamental interface processes important for thin film solar cells. These have been found to be considerably more complex than the well understood interfaces of elemental and III-V semiconductors, which have been extensively investigated since the late 1970's. For II-VI and analogue Cu-chalcopyrite semiconductors the creation of crystallographic defects during interface formation plays a very important role. The high mobility of the Cu atoms in the CIGS also has to be taken into account. Much more systematic studies have to be carried out in order to reach a complete picture of the processes occurring at thin film solar cell interfaces. These experiments require a further development of suitable preparation and deposition techniques in order to account for the full variety of material combinations and preparation processes. Furthermore, the samples need to be investigated with additional techniques. In particular, the combined application of PES and electrical analysis techniques at the same samples seems to be a promising route to understand the correlation between interface properties and solar cell characteristics.

### Acknowledgements

This work has been carried out in close relation to W. Jaegermann, whom I have to thank for the very stimulating atmosphere and the extraordinary experimental facilities. D. Kraft, J. Fritsche, A. Thissen, B. Späth, T. Schulmeyer, R. Hunger and T. Löher have performed or contributed to the experiments presented and discussed in this contribution. T. Mayer and F. Säuberlich have further contributed to the setup of experimental facilities and to the studies of thin film solar cell interfaces at Darmstadt University of Technology. I also have to thank the Zentrum für Sonnenenergie und Wasserstoffforschung (ZSW) and Antec GmbH for their contributions. Our work is funded by the German Government, the Volkswagenstiftung and the German Science foundation.

### References

1. M.A. Contreras, B. Egaas, K. Ramanathan, J. Hiltner, A. Swartzlander, F.A. Hasoon and R. Noufi, *Prog. Photovolt. Res. Appl.* **7**, 311 (1999).
2. T. Dullweber, G. Hanna, U. Rau and H.W. Schock, *Sol. Energy Mat. Sol. Cells* **67**, 145 (2001).
3. D. Schmid, M. Ruckh, F. Grunwald and H.W. Schock, *J. Appl. Phys.* **73**, 2902 (1993).
4. W. Mönch, *Semiconductor Surfaces and Interfaces* (Springer Verlag, Heidelberg, 1993).
5. E.T. Yu, J.O. McCaldin and T.C. McGill, *Band Offsets in Semiconductor Heterojunctions*, in *Solid State Physics* edited by H. Ehrenreich and D. Turnbull (Academic Press, Boston 1992).
6. A. Franciosi and C.G. Van de Walle, *Surf. Sci. Rep.* **25**, 1 (1996).

7. M. Peressi, N. Binggeli and A. Baldereschi, J. Phys. D **31**, 1273 (1998).
8. J. Tersoff, Phys. Rev. Lett. **52**, 465 (1984).
9. S.G. Louie, J.R. Chelikowsky and M.L. Cohen, Phys. Rev. B **15**, 2154 (1977).
10. M. Cardona and N.E. Christensen, Phys. Rev. B **35**, 6182 (1987).
11. W.E. Spicer, Z. Lilienthal-Weber, E. Weber, N. Newman, T. Kendelewicz, R. Cao, C. McCants, R. Mahowald, K. Miyano and I. Lindau, J. Vac. Sci. Technol. B **6**, 1245 (1988).
12. W. Walukiewicz, J. Vac. Sci. Technol. B **5**, 1062 (1987); Phys. Rev. B **37**, 4760 (1988).
13. S. Hüfner, *Photoelectron spectroscopy* (Springer-Verlag, Berlin 1995).
14. D. Briggs and M.P. Seah, *Practical surface analysis by Auger and X-Ray photoelectron spectroscopy* (John Wiley & Sons, New York 1983).
15. F. Capasso and G. Margaritondo, *Heterojunction Band Discontinuities* (North-Holland, Amsterdam, 1987).
16. R. Hunger, T. Schulmeyer, A. Klein, W. Jaegermann, K. Sakurai, A. Yamada, P. Fons, K. Matsubara and S. Niki, Surf. Sci. Lett. (submitted).
17. T. Schulmeyer, R. Kniese, R. Hunger, W. Jaegermann, M. Powalla and A. Klein, Thin Solid Films **451-452**, 420-423 (2004).
18. T. Schulmeyer, R. Hunger, W. Jaegermann, A. Klein, R. Kniese and M. Powalla, Proc. of the 3rd World Conference on Photovoltaic Energy Conversion, Osaka, Japan (2003).
19. A. Niemegeers and M. Burgelman, J. Appl. Phys. **81**, 2881 (1997).
20. D. Kraft, *Präparation und Charakterisierung von Dünnschichtmaterialsystemen für die Rückkontaktbildung bei polykristallinen CdTe-Dünnschichtsolarzellen* (Darmstadt University of Technology, 2004).
21. J.F. McGilp, J. Phys. C **17**, 2249 (1984).
22. J.L. Shaw, R.E. Viturro, L.J. Brillson, D. Kilday, M. K. Kelly and G. Margaritondo, J. Vac. Sci. Technol. A **6**, 2752 (1988).
23. A.K. Wahi, K. Miyano, G.P. Carey, T.T. Chiang, I. Lindau and W.E. Spicer, J. Vac. Sci. Technol. A **8**, 1926 (1990).
24. S.-H. Wei and S. B. Zhang, Phys. Rev. B **66**, 155211 (2002).
25. S.B. Zhang, S.-H. Wei and A. Zunger, J. Appl. Phys. **83**, 3192 (1998).
26. B. Späth et al. (unpublished results).
27. A. Klein and T. Schulmeyer, *Interfaces of Cu-Chalcopyrites*, in *Wide Gap Chalcopyrites*, edited by S. Siebentritt and U. Rau (Springer Verlag, Heidelberg, in press).
28. S.-H. Wei and A. Zunger, J. Appl. Phys. **78**, 3846 (1995); Appl. Phys. Lett. **72**, 2011 (1998).
29. A. Klein, J. Fritzsche, W. Jaegermann, J.H. Schön, C. Kloc and E. Bucher, Appl. Surf. Sci. **166**, 508 (2000).
30. T. Löher, W. Jaegermann and C. Pettenkofer, J. Appl. Phys. **77**, 731 (1995).
31. T. Löher, A. Klein, C. Pettenkofer and W. Jaegermann, J. Appl. Phys. **81**, 7806 (1997).
32. A. Klein and W. Jaegermann, Appl. Phys. Lett. **74**, 2283 (1999).
33. A. Klein, T. Löher, C. Pettenkofer and W. Jaegermann, J. Appl. Phys. **80**, 5039 (1996).
34. A. Klein, Appl. Phys. Lett. **77**, 2009 (2000).
35. Y. Gassenbauer et al. (unpublished results).
36. J. Fritzsche, S. Gunst, A. Thissen, R. Gegenwart, A. Klein and W. Jaegermann, Mater. Res. Soc. Symp. Proc. **668**, H5.1 (2001).

37. F. Säuberlich and A. Klein (submitted to Solid State Ionics)
38. J. Fritzsche, *Halbleitergrenzflächen polykristalliner CdTe Dünnschichtsolarzellen* (Mensch und Buch Verlag, Berlin 2003).

# Electronic Properties of Cu(In,Ga)Se<sub>2</sub> Thin-Film Solar Cells – An Update

Uwe Rau

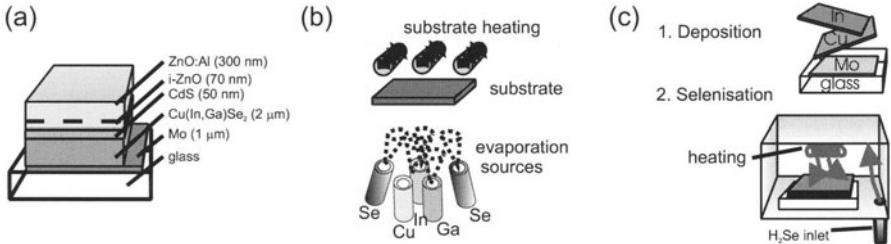
Institut für Physikalische Elektronik, Universität Stuttgart, Germany

**Abstract.** Recent progress in the experimental analysis, modelling, and understanding of ZnO/CdS/Cu(In,Ga)Se<sub>2</sub>-based heterojunction thin-film solar cells is reviewed. Two issues that are inherently critical for the performance of thin-film solar cells are the polycrystallinity of absorber material and the non-lattice matched interface to the CdS buffer layer. The body of recent experimental evidence indicates that Cu(In,Ga)Se<sub>2</sub> copes very well with these challenges. A widening of the band gap at the Cu(In,Ga)Se<sub>2</sub> surface minimizes interface recombination. Grain boundaries in Cu(In,Ga)Se<sub>2</sub> turn out to be electronically almost inactive.

## 1 Introduction

With a power conversion efficiency of 19.2 % for a 0.5 cm<sup>2</sup> laboratory cell [1] and 16.6 % for mini-modules with an area of around 20 cm<sup>2</sup> [2] Cu(In,Ga)Se<sub>2</sub> today is by far the most efficient thin-film solar cell technology. This present status is based on a long history of research and technological development dating back to 1953 when CuInSe<sub>2</sub> was synthesized for the first time [3] and to 1974 when this material was proposed as a photovoltaic material [4]. At present, the start of production at several places in the US, Japan, and Europe provides a new challenge for research on this material.

CuInSe<sub>2</sub> and CuGaSe<sub>2</sub>, the materials that form the alloy Cu(In,Ga)Se<sub>2</sub>, belong to the semiconducting I-III-VI<sub>2</sub> materials family that crystallizes in the tetragonal chalcopyrite structure. The alloy system of the Cu-chalcopyrites Cu(In,Ga,Al)(Se,S)<sub>2</sub> includes a wide range of band gap energies from 1.04 eV in CuInSe<sub>2</sub> up to 2.4 eV in CuGaS<sub>2</sub>, and even 2.7 eV in CuAlS<sub>2</sub>, thus covering most of the visible spectrum. All these compounds have a direct band gap making them suitable for thin-film photovoltaic absorber materials. Many efforts are made to use the full alloy system of Cu-chalcopyrites to produce solar cells, e.g., with a large band gap energy ( $E_g \geq 1.5$  eV) to provide the top cell of an all thin-film tandem structure (for a recent review on wide-gap chalcopyrites see [6]). The present article restricts itself on Cu(In,Ga)Se<sub>2</sub> with a moderate Ga-content  $[Ga]/[Ga]+[In] \approx 0.2\text{--}0.3$  with  $E_g \approx 1.1\text{--}1.2$  eV, the standard material that is used for the to date most efficient cells as well as for the current efforts for commercialization. The article focusses on recent progress in our understanding of issues relevant for photovoltaic applications of Cu(In,Ga)Se<sub>2</sub> solar cells and should be read as an update to a series of earlier review articles [7,8,9].



**Fig. 1.** Layer sequence of the ZnO/CdS/Cu(In,Ga)Se<sub>2</sub> heterojunction solar cell (a). High-quality Cu(In,Ga)Se<sub>2</sub> absorbers can be prepared by two different processes, the co-evaporation of the elements (b) or by a sequential process (c) using the deposition of metallic precursors and a subsequent selenization step

## 2 Device Technology

This section deals with the production methods for ZnO/CdS/Cu(In,Ga)Se<sub>2</sub> heterojunction solar cells. Some emphasis is put on two critical preparation steps, namely the growth of the Cu(In,Ga)Se<sub>2</sub> absorber material, including the role of Na, and on the formation of the CdS/Cu(In,Ga)Se<sub>2</sub> heterojunction. Both preparation steps appear decisive for the final device performance but still are not sufficiently well understood. Figure 1a shows the layer sequence of a ZnO/CdS/Cu(In,Ga)Se<sub>2</sub> heterojunction solar cell together with the two most important preparation routes for the absorber, namely co-evaporation of the elements (Fig. 1b) and the sequential process (Fig. 1c), where metallic precursor layers are deposited by sputtering and compound formation follows by selenization in a second step. In both processes Cu(In,Ga)Se<sub>2</sub> compound formation requires a temperature in the range of 550 °C.

### 2.1 Absorber Preparation and the Role of Na

The preparation of Cu(In,Ga)Se<sub>2</sub>-based solar cells starts with the deposition of the absorber material on a Mo-coated glass substrate (preferably soda-lime glass). The properties of the Mo film and the choice of the glass substrate are of primary importance for the final device quality, because of the importance of Na, which diffuses from the glass through the Mo film into the growing absorber material. Some processes, like the sequential preparation route, use blocking layers between the glass substrate and the Mo film to prevent the out-diffusion of Na. Instead, Na-containing precursors like NaF, Na<sub>2</sub>Se, or Na<sub>2</sub>S are deposited prior to absorber growth to provide a controlled incorporation of Na into the film.

The most obvious effects of Na incorporation are better film morphology and higher conductivity of the films [10]. Furthermore, the incorporation of Na induces beneficial changes in the defect distribution of the absorber films [11,12]. The explanations for the beneficial impact of Na are manifold,

and it is most likely that the incorporation of Na in fact results in a variety of consequences. During film growth, the incorporation of Na leads to the formation of NaSe<sub>x</sub> compounds. This slows down the growth of Cu(In,Ga)Se<sub>2</sub> and could at the same time facilitate the incorporation of Se into the film [13]. Also the widening of the existence range of the  $\alpha$ -(CuInSe<sub>2</sub>)-phase in the phase diagram [14] as well as the larger tolerance to the Cu/(In + Ga) ratio of Na-containing thin films could be explained in this picture. Furthermore, the higher conductivity of Na-containing films could result from the diminished number of compensating V<sub>Se</sub> donors. However, recent experiments using a post-deposition Na incorporation into Na-free Cu(In,Ga)Se<sub>2</sub> films [15] show that there is a beneficial effect of Na *after* film growth as well. This experimental finding is compatible with the model that one major effect of Na is the passivation of grain boundaries [16].

Photovoltaic-grade Cu(In,Ga)Se<sub>2</sub> films have a slightly In-rich overall composition. However, Cu(In,Ga)Se<sub>2</sub> with a Cu-rich composition plays an important role *during* film growth. Films that go through a Cu-rich growth stage have grain sizes in excess of 1  $\mu\text{m}$  whereas films grown under continuous Cu deficiency have much smaller grains. A model for the film growth under Cu-rich compositions comprises the role of Cu<sub>2-y</sub>Se as a flux agent during the growth process of co-evaporated films [17]. For Cu(In,Ga)Se<sub>2</sub> prepared by selenisation, the role of Cu<sub>2-y</sub>Se is similar. Therefore, growth processes for high quality material have to go through a Cu-rich stage but have to end up with a Cu-poor overall composition. The benefit of this final Cu-poor composition stems from a Cu-poor surface layer that forms spontaneously on those films. This so-called surface defect layer (SDL) minimizes recombination at the CdS/Cu(In,Ga)Se<sub>2</sub> heterointerface as will be discussed below in more detail.

## 2.2 Buffer and Window Deposition

Heterojunction formation is most easily achieved by chemical bath deposition (CBD) of a thin CdS film from a solution containing Cd ions and thiourea [18]. The benefit of the CdS layer is manifold:

CdS from CBD provides complete coverage of the rough polycrystalline absorber surface at a film thickness of only 10 nm. The layer provides protection against damage and chemical reactions resulting from the subsequent ZnO deposition process. The chemical bath removes the natural oxide from the film surface [18] and thus re-establishes positively charged surface states and, as a consequence, the natural type inversion at the CdS/Cu(In,Ga)Se<sub>2</sub> interface [19]. The Cd ions, reacting first with the absorber surface, remove elemental Se, possibly by the formation of CdSe. To a certain extent, Cd ions also diffuse into the Cu-poor surface layer of the absorber material [20,21], where they possibly form Cd<sub>Cu</sub> (i.e., Cd on a Cu site) donors, thus providing additional positive charges enhancing the type inversion of the buffer/absorber interface.

Due to the favourable properties of CdS as a heterojunction partner and due to the chemistry of the CBD process, it is difficult to find a replacement. Avoiding CdS as well as the chemical bath step would be advantageous from the production point of view. On the one hand, a toxic material such as CdS requires additional safety regulation, on the other hand, the chemical bath deposition does not comply with the vacuum deposition steps of an in-line module fabrication. Therefore, research and development in this area relates to two issues: (i) the search for alternative materials for a chemical deposition, and (ii) the development of ways to deposit the front electrode without an intermediate step in a chemical bath. Promising materials to replace CdS are In(OH,S), Zn(OH,S), ZnS, ZnSe, and In<sub>2</sub>S<sub>3</sub>. However, all these materials require additional precautions to be taken for the preparation of the absorber surface or front electrode deposition.

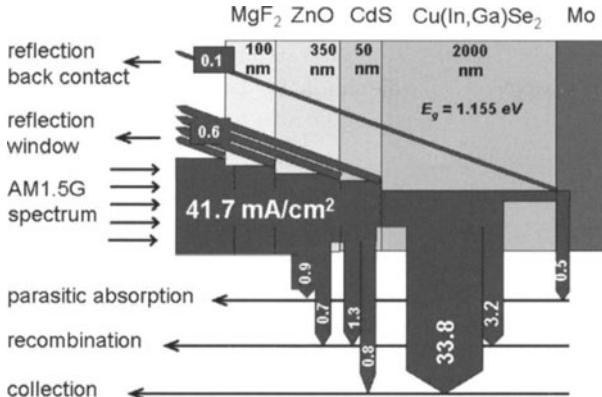
The most commonly used material for the preparation of the front electrode is ZnO doped with B or Al. The first large-area modules produced by ARCO Solar (later Siemens Solar, now Shell Solar Industries) had a ZnO:B window layer deposited by chemical vapour deposition (CVD). Later production facilities at Boeing and EUROCIS use sputtering processes. Present pilot production lines also favour sputtering. As mentioned above, an undoped i-ZnO layer with a thickness of about 50–100 nm is needed at the heterojunction in order to achieve optimum performance.

### 3 Device Physics

This section deals with electronic and optical loss mechanisms that determine the photovoltaic output parameters short circuit current density  $j_{SC}$  and open circuit voltage  $V_{OC}$ . The section will discuss the most important and specific features of a polycrystalline photovoltaic thin-film device, namely the optics of the thin layer stack, the role of the heterointerface for recombination, the defect physics of Cu(In,Ga)Se<sub>2</sub>, and the influence of polycrystallinity on the device performance.

#### 3.1 Short Circuit Current

The short current density that can be obtained from the standard 100 mW cm<sup>-2</sup> solar spectrum (AM1.5) is determined, on the one hand, by *optical losses*, that is, by the fact that photons from a part of the spectrum are either not absorbed in the solar cell or are absorbed without generation of electron-hole pairs. On the other hand, not all photogenerated electron-hole pairs contribute to  $j_{SC}$  because they recombine before they are collected. We denote these as *recombination losses*. Figure 2 illustrates how much from an incoming photon flux from the terrestrial solar spectrum contributes to the final  $j_{SC}$  of a highly efficient Cu(In,Ga)Se<sub>2</sub> solar cell [22] and where the remainder gets lost. The incoming light, i.e., that part of the solar spectrum with photon energy  $E_\gamma$  larger than the band gap energy  $E_g = 1.155$  eV



**Fig. 2.** Diagram of optical and electronic losses for the short circuit current density  $j_{SC}$  after [22]. From the available  $41.7 \text{ mA cm}^{-2}$  only 83 % ( $33.8 \text{ mA cm}^{-2}$ ) from the absorber and  $0.8 \text{ mA cm}^{-2}$  from the buffer layer contribute to the real  $j_{SC}$  of a good Cu(In,Ga)Se<sub>2</sub> solar cell

of the specific absorber would correspond to a (maximum possible)  $j_{SC}$  of  $41.7 \text{ mA cm}^{-2}$ . By reflection at the surface and at the interfaces between the MgF<sub>2</sub> anti-reflective coating, the ZnO window, the CdS buffer, and the Cu(In,Ga)Se<sub>2</sub> absorber we loose already  $0.6 \text{ mA cm}^{-2}$ . An additional loss of  $0.1 \text{ mA cm}^{-2}$  stems from photons that are re-emitted after one pass through the absorber. Free carrier absorption in the highly doped ZnO window layer costs  $0.9 \text{ mA cm}^{-2}$  and absorption at the absorber/Mo interface  $0.5 \text{ mA cm}^{-2}$ . Thus, the sum of all optical losses amounts to  $2.1 \text{ mA cm}^{-2}$ . Additional optical losses due to the grid shadowing are not considered.

Absorption of photons with energies  $E_\gamma > E_g(\text{ZnO}) = 3.2 \text{ eV}$  in the window layer amounts to about  $0.7 \text{ mA cm}^{-2}$ . Another portion of the solar light is absorbed in the CdS buffer layer ( $E_\gamma > E_g(\text{CdS}) > 2.4 \text{ eV}$ ). A part of these photons with  $3.2 \text{ eV} > E_\gamma > 2.4 \text{ eV}$  contributes to  $j_{SC}$ , because the thin CdS layer does not absorb all those photons and only a part of the electron-hole pairs created in the buffer layer recombines [22]. The loss due to incomplete collection in the absorber amounts to  $3.2 \text{ mA cm}^{-2}$ . The above analysis shows that there is still considerable scope of improving  $j_{SC}$  in Cu(In,Ga)Se<sub>2</sub> solar cells, by improving carrier collection from the absorber and minimization of buffer absorption. Thus, a possible gain of about  $4 \text{ mA cm}^{-2}$  in  $j_{SC}$  could enhance the efficiency by more than 10 % relative.

### 3.2 Open Circuit Voltage

The band diagram in Fig. 3 shows the possible recombination paths for the photogenerated charge carriers in the Cu(In,Ga)Se<sub>2</sub> absorber. The figure consideres recombination at the back surface of the absorber (A') and in the

neutral bulk (A), recombination in the space charge region (B), and recombination at the buffer/absorber interface (C). Note that due to the presence of high electric fields in the junction region, the latter two mechanisms may be enhanced by tunnelling [23,24]. The basic equations for the recombination processes (A-C) can be, e.g., found in Ref. [25]. Notably, all recombination current densities  $j_R$  may be written in the form of a diode law

$$j_R = j_0 \left\{ \exp \left( \frac{qV}{n_{id}kT} \right) - 1 \right\} = j_{00} \exp \left( \frac{-E_a}{n_{id}kT} \right) \left\{ \exp \left( \frac{qV}{n_{id}kT} \right) - 1 \right\} \quad (1)$$

where  $V$  is the applied voltage,  $n_{id}$  the diode ideality factor, and  $kT/q$  the thermal voltage. The saturation current density  $j_0$  in general is thermally activated with energy  $E_a$  and the pre-factor  $j_{00}$  is only weakly temperature-dependent. The activation energy for back surface recombination, neutral zone, and space charge recombination is the absorber band gap energy  $E_g$  whereas in case of interface recombination equals the barrier  $\phi_{bp}$  that hinders the holes from the absorber to come to the buffer/absorber interface (cf. Fig. 3). In the simplest cases, the diode ideality factor is unity for back surface and neutral zone recombination as well as for recombination at the buffer/absorber interface, whereas for space charge recombination  $1 \leq n_{id} \leq 2$ . Equalizing short circuit current and recombination current density yields

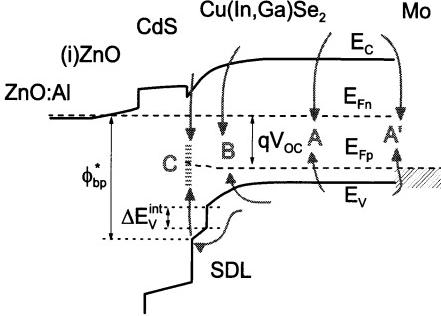
$$V_{OC} = \frac{E_a}{q} - \frac{n_{id}kT}{q} \ln \left( \frac{j_{00}}{j_{SC}} \right) \quad (2)$$

Note that (2) is strictly valid only if a *single* mechanism dominates recombination. For a large series of Cu(In,Ga)(Se,S)<sub>2</sub> devices the dominant recombination mechanism was identified to be related to the bulk of the absorber material [26] as long as these devices have an overall Cu-poor composition. Similarly, a series of Cu(In<sub>1-x</sub>Ga<sub>x</sub>)Se<sub>2</sub> devices with a Ga-content  $x$  varying between  $x = 0$  and  $x = 1$  exhibited dominant bulk recombination except for  $x = 1$  [27]. However, a detailed study on CuGaSe<sub>2</sub> devices prepared under a variety of conditions showed that only the best CuGaSe<sub>2</sub> devices were dominated by bulk recombination whereas less optimized preparations led to devices that are dominated by interface recombination [28].

The absence of interface recombination in almost all Cu(In<sub>1-x</sub>Ga<sub>x</sub>)Se<sub>2</sub> devices with Cu-poor composition and the dominance of this recombination path for devices that stem from Cu-rich absorbers [26,29] indicates that the Cu-poor surface defect layer (SDL) might play an important role in suppressing interface recombination as will be discussed in the following.

### 3.3 Surface Band Gap Widening

The band diagram in Fig. 3 contains an important features that appear important for the minimization of recombination losses in Cu(In,Ga)Se<sub>2</sub> solar cells. The first one is the presence of a considerably widened band gap at the



**Fig. 3.** Band diagram of the ZnO/CdS/Cu(In,Ga)Se<sub>2</sub> heterojunction under applied voltage  $V = V_{OC}$ . The internal valence band offset  $\Delta E_V^{int}$  between the surface defect layer (SDL) and the bulk of the Cu(In,Ga)Se<sub>2</sub> absorber enhances the hole barrier  $\phi_{bp}^*$  at the CdS/Cu(In,Ga)Se<sub>2</sub> interface. Note that the interface between the SDL and the bulk absorber is drawn here as an abrupt transition only for clarity. In reality, this transition as well as the band offset  $\Delta E_V^{int}$  may occur more gradually

surface of the Cu(In,Ga)Se<sub>2</sub> absorber film. This surface band gap widening leads to a valence band offset  $\Delta E_V^{int}$  between the absorber and the SDL. This internal offset implies that the recombination barrier  $\phi_{bp}$  is increased via  $\phi_{bp}^* = \phi_{bp} + \Delta E_V^{int}$ . Following Fig. 3 the activation energy  $E_a$  for interface recombination is then  $E_a = \phi_{bp}^*$  and, as a consequence,  $V_{OC}^{ifr}$  for interface recombination reads

$$V_{OC}^{ifr} = \frac{\phi_{bp} + \Delta E_V^{int}}{q} - \frac{n_{id}kT}{q} \ln \left( \frac{j_{00}}{j_{SC}} \right). \quad (3)$$

The internal band offset directly adds to the open circuit voltage that is allowed if only interface recombination were present. A more detailed mathematical treatment of the ZnO/CdS/Cu(In,Ga)Se<sub>2</sub> heterojunction including the consequence of the surface band gap widening on interface recombination is given in Ref. [30]. Eventually, because of this surface band gap widening most Cu-poor devices are *not* limited by the interface. Instead, recombination in the bulk is dominant [30].

First experimental evidence for this very special feature of polycrystalline Cu(In,Ga)Se<sub>2</sub> films stems from 1993 [31]. There it was found that the surface composition of Cu-poor CuInSe<sub>2</sub> as well as that of Cu(In,Ga)Se<sub>2</sub> films, corresponds to a surface composition of  $[Ga + In]/[Ga + In + Cu]$  of about 0.75 for a range of bulk compositions of  $0.5 < [Ga + In]/[Ga + In + Cu] < 0.75$ . In contrast, in Cu-rich films ( $[Ga + In]/[Ga + In + Cu] < 0.5$ ) the surface composition follows the bulk composition [31,32]. This observations has led to the assumption that a phase segregation of Cu(In,Ga)<sub>3</sub>Se<sub>5</sub>, the so-called Ordered Defect Compound (ODC), occurs at the surface of Cu-poor films. However, the existence of a *separate phase* on top of polycrystalline Cu(In,Ga)Se<sub>2</sub> thin films has never been confirmed. To the contrary, recent high-resolution electron microscopy analysis using convergent-beam electron diffraction [33] practically rules out such a separate phase. Therefore, the

less specific denotation 'surface defect layer (SDL)' [34] was chosen instead of ODC. However, recent results from grazing incidence X-ray diffraction [35] unveiled a Cu-depleted region on the surface of polycrystalline Cu(In,Ga)Se<sub>2</sub> films. These layers had thicknesses between 5 and 60 nm, dependent on the overall Cu-content of the film.

The Photo Electron Spectroscopy experiments of Schmid et al. [31,32] have shown that the Fermi level at the surface of as-grown CuInSe<sub>2</sub> films lies above the valence band-edge by about 1.1 eV. This energy difference is larger than the band-gap energy of the bulk. Therefore, the band gap at the surface of the film must be larger than that of the bulk. A recent direct measurement of the surface band gap of polycrystalline CuInSe<sub>2</sub> [36] proofed that the band gap energy  $E_g^{surf}$  at the surface of the film is about  $E_g^{surf} \approx 1.4$  eV, i.e., more than 0.3 eV larger than  $E_g^{bulk} \approx 1.04$  eV. More evidence for the surface band gap widening stems from recent highly resolved cathodoluminescence experiments [37] that show emission lines with higher photon energies in regions that could be identified with the SDL.

Despite of some controversy on the specific nature of the SDL (see, e.g., Refs. [14,34,38]), there appears more and more an agreement that the surface band gap widening going along with the SDL is the primary and unique feature of the Cu(In,Ga)Se<sub>2</sub> surface. Numerical device simulations unveil that suppression of interface recombination by the surface band gap widening is active regardless of specific assumptions on the bulk properties of the SDL [39]. Even an increased amount of structural defects as observed in the SDL [40] is compatible with a high power conversion efficiency [39]. However, maintaining the beneficial properties of the SDL during heterojunction formation is a critical issue as suggested by the aforementioned difficulties in replacing the CBD-CdS by alternative materials or other deposition methods.

### 3.4 Defects

The role of defects in the ternary compounds CuInSe<sub>2</sub>, CuGaSe<sub>2</sub>, CuInS<sub>2</sub>, and their alloys, is of special importance because of the large number of possible intrinsic defects and their role for doping and as deep recombination centers. The features that are somewhat special to those Cu-chalcopyrites are the ability to dope these compounds with native defects, their tolerance to large off-stoichiometries, and the electrically neutral nature of structural defects [41].

Let us now concentrate on the defects experimentally detected in photovoltaic grade (and thus In-rich) polycrystalline films. In-rich Cu(In,Ga)Se<sub>2</sub> is in general highly compensated, with a net acceptor concentration of the order of  $10^{15}$  to  $10^{16}$  cm<sup>-3</sup>. The shallow acceptor level  $V_{Cu}$  is assumed to be the main dopant in this material. As compensating donors the Se-vacancy  $V_{Se}$  as well as the double donor In<sub>Cu</sub> are considered. Recent combined Hall and photoluminescence (PL) measurements on CuGaSe<sub>2</sub> [42] and CuInSe<sub>2</sub> [43] epitaxial films unveiled two shallow acceptor states and one compensating

donor in both materials. The expected strong tendency of Cu-chalcopyrites for self-compensation was experimentally verified recently by PL analysis of CuGaSe<sub>2</sub> epitaxial and polycrystalline films [44]. Self-compensation plays an especially important role for the photovoltaic grade Cu-poor Cu(In,Ga)Se<sub>2</sub>: The high degree of compensation leads to potential fluctuations that affect radiative recombination [45] as well as the maximum possible efficiency [46].

The most prominent *deep defect* is an acceptor level at about 270–300 meV above the valence band. The concentration of this defect results is related to the open-circuit voltage of the device [47,48]. Upon investigating defect energies in the entire Cu(In,Ga)(Se,S)<sub>2</sub> alloy system, Turcu et al. [49] found that the energy distance between this defect and the valence band maximum remains constant when alloying CuInSe<sub>2</sub> with Ga, whereas the energy distance increases under S/Se alloying. Recently, transient photocapacitance studies by Heath et al. [50] unveiled an additional defect state in Cu(In,Ga)Se<sub>2</sub> at about 0.8 eV from the valence band. Again, the defect energy is independent from the Ga content in the alloy.

### 3.5 Polycrystallinity

Despite of the fact that the grain size of the Cu(In,Ga)Se<sub>2</sub> absorber material is relatively small (with around 2  $\mu\text{m}$  about 1000 times smaller than in multi-crystalline Si), there are only few experimental investigations on grain boundaries (GB) in Cu(In,Ga)Se<sub>2</sub> or CuGaSe<sub>2</sub>. Temperature dependent Hall-effect measurements on polycrystalline CuGaSe<sub>2</sub> [51] unveiled a band bending of around 100 meV at the GB of this material. This finding fits well to the conclusions drawn from Kelvin-probe force microscopy (KPFM) experiments performed on the same material [52]. For Cu(In,Ga)Se<sub>2</sub> Hall measurements at room temperature unveiled a GB barrier height of roughly 200 meV [53]. A GB band bending of only 100–200 meV is already indicative for a low electronic activity. Recent numerical simulations of polycrystalline Cu(In,Ga)Se<sub>2</sub> solar cells [54] showed that only with a low recombination velocity  $S_{gb} < 10^3 \text{ cm/s}$ , efficiencies above 19 % are possible when using a material of grain size  $g \approx 2 \mu\text{m}$ . The reasons for the low recombination activity of GBs in Cu(In,Ga)Se<sub>2</sub> are still under debate. One might think about a self-passivation [55] of the grain surface by a band gap widening similar to that of the SDL at the film surface. However, GB activity appears not to be independent from the film preparation method. A recent complementary study on CIGS using KPFM and cathodoluminescence (CL) analysis [56] shows that differently oriented Cu(In,Ga)Se<sub>2</sub> films exhibit different electronic activity of their GBs. This result holds for the band bending at the GB as well as for the recombination rate both being larger in (112)-oriented compared to (220/204)-oriented films. This finding fits well to the fact that Cu(In,Ga)Se<sub>2</sub> films with (220/204)-texture are those that deliver the highest power conversion efficiency [57]. Other CL studies [58] unveiled that Na incorporation appears necessary for a low recombination rate at GBs. This result gives

strong support for the proposition of Kronik et al. [16] that Na-catalyzed oxygenation of dangling bonds leads to the passivation of GBs.

Whether or not the polycrystalline nature of Cu(In,Ga)Se<sub>2</sub> actually is the limiting factor for its photovoltaic performance is not yet clarified. However, looking more generally at the material's electronic inhomogeneity might be a worthwhile effort [59]. E.g., spatially resolved photoluminescence measurements unveiled that the split of the quasi-Fermi levels, i.e. the local  $V_{OC}$ , in Cu(In,Ga)Se<sub>2</sub> films exhibits spatial variations with an amplitude of 30 mV [60]. If such variations result from spatial fluctuations of the fundamental band gap, these inhomogeneities would lead to a loss of the overall photovoltaic efficiency of 2 % absolute [46].

## 4 Conclusions

Considering two inherent challenges of a polycrystalline thin-film solar technology, namely (i) the presence of a non-lattice matched heterointerface between the photovoltaic absorber material and the heterojunction partner (the buffer/window combination) and (ii) the polycrystalline nature of the absorber, Cu(In,Ga)Se<sub>2</sub>-based solar cells, from a physical point of view, appear to be the almost best imaginable solution. Recombination at the heterointerface is suppressed by the spontaneous formation of the surface band gap widening, the electronic activity of GBs is low, and, most importantly, it can be manipulated. Further improvements in our understanding of film growth and junction formation mechanisms will be of help to enhance device efficiency and to make module production more reliable.

## Acknowledgements

The author thanks G. Hanna, J. Mattheis, K. Orgassa, J. Rostan, K. Taretto, M. Turcu, F. Pfisterer, and H. W. Schock for fruitful collaboration and discussion, as well as J. H. Werner for continuous support and many controversial arguments. Some of the work reviewed in this article has been supported by the German Ministry for Research and Education.

## References

1. K. Ramanathan, A. Contreras, C. L. Perkins, S. Asher, F. S. Hasoon, J. Keane, D. Young, M. Romero, W. Metzger, R. Noufi, J. Ward, A. Duda, *Progr. Photovolt.: Res. Appl.* **11**, 225-230 (2003).
2. J. Kessler, M. Bodegård, J. Hedström, L. Stolt: in *Proc. 16th European Photovoltaic Solar Energy Conference*, H. Scheer, B. McNelis, W. Palz, H. A. Ossenbrink, and P. Helm (Eds.) (James & James, London, 2000) p. 2057-2061.
3. H. Hahn, G. Frank, W. Klingler, A. Meyer, G. Störger, *Z. anorg. u. allg. Chemie* **271**, 153 (1953).

4. S. Wagner, J. L. Shay, P. Migliorato, H. M. Kasper, *Appl. Phys. Lett.* **25**, 434 (1974).
5. J. L. Shay, J. H. Wernick, *Ternary Chalcopyrite Semiconductors: Growth, Electronic Properties, and Applications* (Pergamon Press, Oxford, 1975).
6. S. Siebentritt, *Thin Solid Films* **403-404**, 1 (2002).
7. H. W. Schock, *Festkörperprobleme – Adv. Sol. St. Phys.* **34**, 147 (1995).
8. U. Rau, H. W. Schock, *Appl. Phys. A* **69**, 131 (1999).
9. B. J. Stanberry, *Crit. Rev. Sol. St.* **27**, 73 (2002).
10. M. Ruckh, D. Schmid, M. Kaiser, R. Schäffler, T. Walter, H. W. Schock, in: Proc. 1st. World Conf. on Photovoltaic Solar Energy Conversion (IEEE Press, New York, 1994) p. 156.
11. B. M. Keyes, F. Hasoon, P. Dippo, A. Balcioglu, F. Aboulfotuh, in: Conf. Record 26th. IEEE Photovoltaic Specialists Conf. (IEEE Press, New York, 1997) p. 479.
12. U. Rau, M. Schmitt, F. Engelhardt, O. Seifert, J. Parisi, W. Riedl, J. Rimmasch, F. Karg, *Sol. St. Commun.* **107**, 59 (1998).
13. D. Braunger, D. Hariskos, G. Bilger, U. Rau, H. W. Schock, *Thin Solid Films* **361-362**, 161 (2000).
14. R. Herberholz, U. Rau, H. W. Schock, T. Hallboom, T. Gödecke, F. Ernst, C. Beilharz, K. W. Benz, D. Cahen: *Eur. Phys. J., Appl. Phys.* **6**, 131 (1999).
15. D. Rudmann, A. F. Cunha, M. Kaelin, F. Kurdesau, H. Zogg, A. N. Tiwari, G. Bilger, *Appl. Phys. Lett.* **8**, 1129 (2004).
16. L. Kronik, D. Cahen, H. W. Schock, *Adv. Mat.* **10**, 31 (1998).
17. R. Klenk, T. Walter, H. W. Schock, D. Cahen, *Adv. Mat.* **5**, 114 (1993).
18. J. Kessler, K. O. Velthaus, M. Ruckh, R. Laichinger, H. W. Schock, D. Lincot, R. Ortega, J. Vedel, in: Proc. 6th. Int. Photovoltaic Solar Energy Conf. (Oxford IBH Publishing, New Delhi, India, 1992) p. 1005.
19. U. Rau, D. Braunger, R. Herberholz, H. W. Schock, J.-F. Guillemoles, L. Kronik, D. Cahen, *J. Appl. Phys.* **86**, 497 (1999).
20. K. Ramanathan, H. Wiesner, S. Asher, D. Niles, R. N. Bhattacharya, J. Keane, M. A. Contreras, R. Noufi, in: Proc. 2nd World Conf. on Photovoltaic Solar Energy Conversion (E. C. Joint Res. Centre, Luxembourg, 1998) p. 477.
21. T. Wada, S. Hayashi, Y. Hashimoto, S Nishiwaki, T. Sato, M. Nishitina, in: Proc. 2nd World Conf. on Photovoltaic Solar Energy Conversion (E. C. Joint Res. Centre, Luxembourg, 1998) p. 403.
22. K. Orgassa, Coherent Optical Analysis of the ZnO/CdS/Cu(In,Ga)Se<sub>2</sub> Thin Film Solar Cell. Ph.D. Thesis, Universität Stuttgart, Germany (2004).
23. U. Rau, *Appl. Phys. Lett.* **74**, 111 (1999).
24. U. Rau, A. Jasenek, H. W. Schock, F. Engelhardt, T. Meyer, *Thin Solid Films* **361-362**, 298 (2000).
25. R. H. Bube, *Photoelectronic Properties of Semiconductors* (Cambridge University Press, Cambridge, 1992).
26. M. Turcu, O. Pakma, U. Rau, *Appl. Phys. Lett.* **80**, 2598 (2002).
27. J. Reiβ, J. Malmström, A. Werner, I. Hengel, R. Klenk, M. C. Lux-Steiner, *Mat. Res. Soc. Symp. Proc.* **668**, H9.4.1-H9.4.6 (2001).
28. V. Nadenau, U. Rau, A. Jasenek, H. W. Schock, *J. Appl. Phys.* **87**, 584 (2000).
29. R. Klenk, *Thin Solid Films* **387**, 135 (2001).
30. M. Turcu, U. Rau, *J. Phys. Chem. Solids* **64**, 1591 (2003).
31. D. Schmid, M. Ruckh, F. Grunwald, H. W. Schock, *J. Appl. Phys.* **73**, 2902 (1993).

32. D. Schmid, M. Ruckh, H. W. Schock, *Appl. Surf. Sci.* **103**, 409 (1996).
33. Y. Yan, K. M. Jones, J. Abushama, M. Young, S. Asher, M. M. Al-Jassim, R. Noufi, *Appl. Phys. Lett.* **81**, 1008 (2002).
34. A. Niemeegers, M. Burgelman, R. Herberholz, U. Rau, D. Hariskos, H. W. Schock, *Prog. Photovolt. Res. Appl.* **6**, 407 (1998).
35. I. M. Kötschau, H. W. Schock, *J. Phys. Chem. Solids* **64**, 1559 (2003).
36. M. Morkel, L. Weinhardt, B. Lohmüller, C. Heske, E. Umbach, W. Riedl, S. Zweigart, F. Karg, *Appl. Phys. Lett.* **79**, 4482 (2001).
37. M. J. Romero, K. M. Jones, J. AbuShama, Y. Yan, M. M. Al-Jassim, R. Noufi, *Appl. Phys. Lett.* **83**, 4731 (2003).
38. C.-S. Jiang, F. S. Hasoon, H. R. Moutinho, H. A. Al-Thani, M. J. Romero, M. M. Al-Jassim, *Appl. Phys. Lett.* **82**, 127 (2003).
39. U. Rau, M. Turcu, *Mater. Res. Soc. Symp. Proc.* **763**, B8.8.1-6 (2003).
40. F. S. Hasoon, Y. Yan, H. Althani, K. M. Jones, H. R. Moutinho, J. Alleman, M. M. Al-Jassim, R. Noufi, *Thin Solid Films* **387**, 1 (2001).
41. S. B. Zhang, S. H. Wei, A. Zunger, H. Katayama-Yoshida, *Phys. Rev B* **57**, 9642 (1998).
42. A. Bauknecht, S. Siebentritt, J. Albert, M. Ch. Lux-Steiner, *J. Appl. Phys.* **89**, 4391 (2001).
43. N. Rega, S. Siebentritt, I. E. Beckers, J. Beckmann, J. Albert, M. Ch. Lux-Steiner, *Thin Solid Films* **431**, 186 (2003).
44. S. Schuler, S. Siebentritt, S. Nishiwaki, N. Rega, J. Beckmann, S. Brehme, M. Ch. Lux-Steiner, *Phys. Rev. B* **69**, 045210 (2004).
45. I. Dirnstorfer, Mt. Wagner, D. M. Hoffmann, M. D. Lampert, F. Karg, B. K. Meyer, *Phys. Stat. Sol. (a)* **168**, 163 (1998).
46. U. Rau, J. H. Werner, *Appl. Phys. Lett.* **84**, 3735 (2004).
47. G. Hanna, A. Jasenek, U. Rau, H. W. Schock, *Phys. Stat. Sol. A* **179**, R7 (2000).
48. U. Rau, M. Schmidt, A. Jasenek, G. Hanna, H. W. Schock, *Sol. Energy Mater. Sol. Cells* **67**, 137 (2001).
49. M. Turcu, I. M. Kötschau, U. Rau, J. Appl. Phys. **91**, 1391 (2002).
50. J. T. Heath, J. D. Cohen, W. N. Shafarman, D. X. Liao, A. A. Rockett, *Appl. Phys. Lett.* **80**, 4540 (2002).
51. S. Siebentritt, S. Schuler, *J. Phys. Chem. Solids* **64**, 1621 (2003).
52. S. Sadewasser, T. Glatzel, S. Schuler, S. Nishiwaki, R. Kaigawa, M. Ch. Lux-Steiner, *Thin Solid Films* **431-432**, 257 (2003).
53. T. Meyer, F. Engelhardt, J. Parisi, U. Rau, J. Appl. Phys. **91**, 5093 (2002).
54. K. Taretto, U. Rau, J. H. Werner, *Thin Solid Films* (in press).
55. C. Persson, A. Zunger, *Phys. Rev. Lett.* **91**, 266401 (2003).
56. G. Hanna, Determination and Influence of Na Supply and Se Flux during Growth of Cu<sub>x</sub>(In,Ga)Se<sub>2</sub> Thin Films. Ph.D. Thesis, Universität Stuttgart, Germany (2004).
57. M. Contreras, B. Egaas, K. Ramanathan, J. Hiltner, A. Swartzlander, F. Hasoon, R. Noufi, *Prog. Photovolt. Res. Appl.* **7**, 311 (1999).
58. N. Ott, G. Hanna, U. Rau, J. H. Werner, H. P. Strunk, *J. Phys.-Condensed Matter* **16**, S85 (2004).
59. J. H. Werner, J. Mattheis, U. Rau, in: *The Path to Ultra High Efficient Photovoltaics*, A. Jaeger-Waldau (Ed.), (JCR, Ispra, Italy, 2004) in press.
60. K. Bothe, G. H. Bauer, T. Unold, *Thin Solid Films* **403**, 453 (2002).

# Progress in Manufacturable High-Efficiency Silicon Solar Cells

Rudolf Hezel

Institut für Solarenergieforschung Hameln/Emmerthal (ISFH),

31860 Emmerthal, Germany

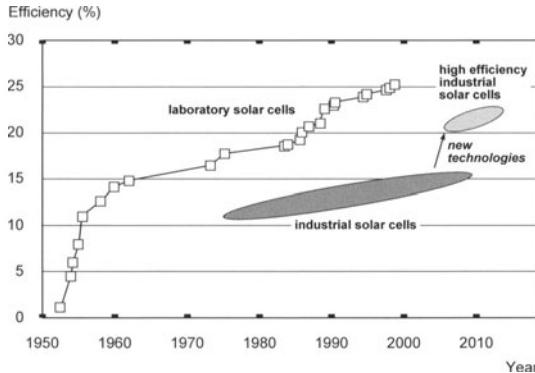
[hezel@isfh.de](mailto:hezel@isfh.de)

**Abstract.** As an important step towards making photovoltaics an economic source of energy, the efficiency of industrial crystalline silicon solar cells has to be drastically increased. In this way not only more power is gained with less silicon, but simultaneously also all other area-related costs of the PV system are reduced. It's of great importance to bridge the efficiency gap existing between the present industrial solar cells with values of 14% to 16% and the sophisticated laboratory cells with a record efficiency close to 25%. Recently new technologies are emerging or existing ones being upgraded aiming for the ambitious goal to manufacture silicon solar cells with efficiencies above 20% cost effectively in an industrial environment. Five different promising approaches of high efficiency silicon solar cells are presented in this paper, including both front and back contacted as well as bifacially sensitive devices. The choice of the silicon substrate material is also discussed.

## 1 Introduction

One of the most important obstacles for a significant global utilization of solar electricity are the high costs currently associated with photovoltaic power supply. Higher production volumes are a major key driver in cost reduction, however innovative processing sequences, combining very high solar cell efficiencies with simple and cost-effective fabrication techniques are needed to become competitive with conventional energy sources. The energy conversion efficiency has a great impact on the costs of a PV system. With a higher efficiency less cell and module area is required to achieve the same output power. Consequently all area-related costs are reduced, which make up more than 70% of the costs of a PV-system. These include costs of the silicon feedstock, wafering, cell processing, module fabrication, installation and maintenance [1]. Already the silicon wafer is contributing about 50% to the module costs. Gaining more power with less silicon is thus one of the most important consequences of an increase in efficiency. Using thinner wafers (100–150  $\mu\text{m}$  instead of the present 300  $\mu\text{m}$ ) as well as making the solar cell bifacially sensitive are further means to reduce the cost of solar electricity.

As to the general situation of crystalline silicon solar cell efficiencies, since the first introduction in 1954 considerable progress has been achieved with laboratory cells characterized by high process complexity. As can be seen

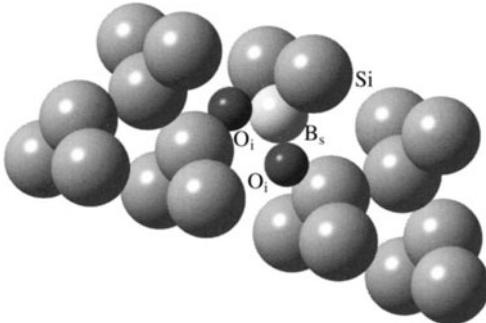


**Fig. 1.** Development of Crystalline Silicon Solar Cell Efficiencies in the laboratory and in industry

in Fig. 1, record efficiencies up to 24.7% are obtained for lab cells with an area far below the minimum industrial size of  $100\text{ cm}^2$  [2,3]. An efficiency of 28.8% was determined as the theoretical limit of a silicon solar cell of  $80\text{ }\mu\text{m}$  thickness and with Lambertian light trapping [4]. The efficiencies of industrial solar cells are presently in the range of 13–16% including both monocrystalline Cz-grown and cast multicrystalline silicon as well as ribbon grown silicon. Despite rapid growth of the manufacturing volume occurred in the past, accompanied by a significant drop in the module selling price, it is indispensable for the economy of PV solar electricity to open up a new high efficiency area above 20% in the near future with industrial crystalline silicon solar cells fabricated by novel cost-effective technologies (see Fig. 1). In particular, processing should be simple, mask- and photolithography-free to be suitable for large-scale production. It should be an important long term goal to bridge the wide efficiency gap presently existing between highly sophisticated small-area laboratory cells and large-area commercial solar cells. In order to obtain high efficiencies both optical and electronic properties of the solar cells have to be optimized. This includes surface texturing and antireflection coating, low grid shadowing, light trapping, silicon with high bulk carrier lifetime, shallow emitter, small contact area and contact passivation, silicon front and back surface passivation etc. [4]. In this paper promising candidates regarded to meet the requirements for mass production of high efficiency silicon solar cells are presented [5].

## 2 Silicon Substrate Options

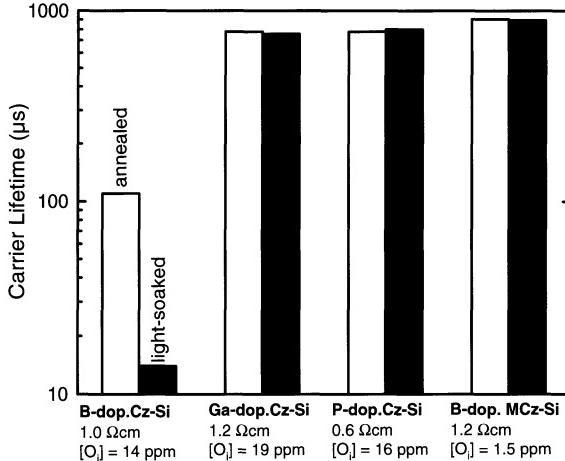
High minority carrier lifetime of the silicon substrate is a prerequisite for high efficiency solar cells. For laboratory devices and for special applications such as race cars or high altitude airplanes float zone (FZ) silicon as the highest quality single crystalline material was used in order to reach record



**Fig. 2.** Possible configuration of the metastable  $B_sO_{2i}$  complex in the silicon host lattice of B-doped Cz-Si (taken from [11])

cell efficiencies. However, for low cost manufacturable solar cells with efficiencies of about 20% Czochralski-grown (Cz) silicon should be preferred. Boron-doped Cz-Si has a market share of about 40% of the present world solar cell production resulting in efficiencies up to 16%. However, a serious problem of solar cells manufactured on Cz-Si is that their initial efficiency degrades under illumination until a stable performance level is reached. For high efficiency cells a degradation by up to 10% relative has been reported, which is presently the main obstacle for making Cz-Si a perfect high-efficiency solar cell material [6]. This degradation effect, already known for 30 years, is due to the activation of a specific metastable defect, which is correlated with the boron and the oxygen concentration in the material [7]. Large concentrations of oxygen are virtually unavoidable in Cz-Si due to the partial dissolution of the silicon crucible during the crystal growth process. Based on a variety of experiments, a new defect reaction model as shown in Fig. 2 was recently developed [8]. In this model, fast-diffusing oxygen dimers  $O_{2i}$  are captured by substitutional boron  $B_s$  to form a  $B_s-O_{2i}$  complex, acting as highly effective recombination center. Due to the fact that the tetrahedral covalent radius of the  $B_s$  atom is 25% smaller than that of the Si host atom, the  $O_{2i}$  is preferentially accommodated in the vicinity of a  $B_s$  atom.

Due to the direct correlation of the magnitude of degradation with the boron and the oxygen concentration in a crystalline Si wafer several methods for elimination of the lifetime degradation in Cz-Si solar cells were proposed [7]. The two most promising approaches are (i) replacement of B with another dopant element, such as Ga or P, and (ii) reduction of the oxygen concentration in the Cz material. The latter can be achieved by damping the melt flows with magnetic fields, resulting in the so-called magnetic-field assisted Cz (MCz) silicon with oxygen concentrations below 1 ppm. Fig. 3 shows the typical behavior of the different materials: while the carrier lifetime of conventional B-doped Cz-Si degrades under illumination, Ga-doped Cz-Si of similar doping concentration has a stable lifetime on a much higher level, comparable to that of B-doped FZ-Si, even if the interstitial oxygen concentration is as high as in the case of the B-doped Cz-Si. Referring to



**Fig. 3.** Measured carrier lifetime of B-, Ga- and P-doped Cz-Si and B-doped MCz-Si before and after light soaking (taken from [11])

the above mentioned oxygen-dimer model, this behavior is attributed to the larger tetrahedral covalent radius of Ga compared to Si, making it less likely to accommodate the oxygen dimer. Due to the two orders of magnitude lower segregation coefficient of gallium compared to boron in Cz-Si, the Ga-doped Cz-Si crystals exhibit a considerably higher variation in resistivity along their growth axis. However, as demonstrated in Fig. 7 for OECO solar cells, these variations are well tolerable [9]. High efficiency solar cell processes were applied to the alternative Cz materials at different institutes and stable efficiencies well above 20% were obtained on Ga-doped Cz-Si, B-doped MCz-Si and P-doped n-type Cz-Si, confirming the lifetime results shown in Fig. 3 [10,11].

Other approaches to at least partially reduce the harmful effects of lifetime degradation on cell performance are using thinner substrates [12] and proper cell processing sequences [11]. Only recently a novel FZ-Si material was introduced for which, due to its high and stable carrier lifetime, high solar cell efficiencies could already be demonstrated [13]. It was reported that the high production costs of FZ-Si, caused by the fact that the starting material has to be in the form of an almost perfectly shaped polycrystalline rod, could be drastically reduced [13].

### 3 High Efficiency Silicon Solar Cells

In the following, five different devices for the high efficiency area are presented including the back contacted Point Contact solar cell, the HIT cell, the Buried Contact cell, the front contacted Standard OECO cell as well as the bifacially sensitive back contacted BACK OECO solar cell.

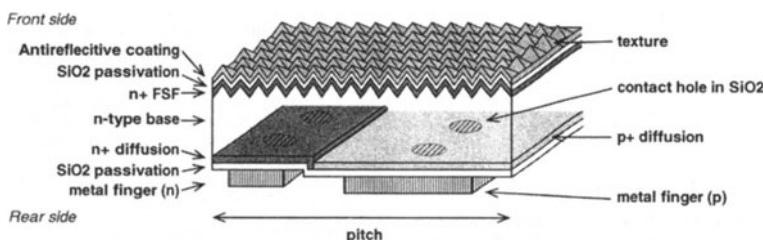
### 3.1 The Point-Contact Solar Cell

The novel A-300 high efficiency cell of Sun Power is a simplified version of the highly sophisticated point contact solar cell designed for concentrator applications and successfully applied for solar cars and solar airplanes [14]. Originally several photolithography steps were applied to define the rear features (i.e. the diffusions, contact openings and fingers), so that one-sun efficiencies approaching 23% could be achieved. In order to make processing suitable for mass production, a novel screen printing technology for masking purposes was developed [5,15]. The device is schematically depicted in Fig. 4.

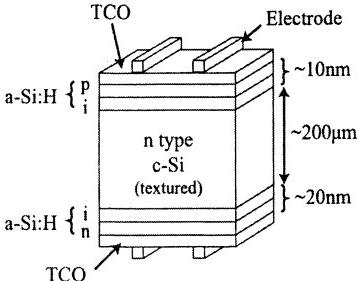
Both contact schemes are placed on the rear side. This is advantageous due to (i) improvement of efficiency since shading losses by the front grid are avoided (ii) reduction of module fabrication costs and (iii) better visual appearance of the PV modules. However, the substrate quality has to be high in terms of minority carrier diffusion length, since the photo-generated charge carriers have to diffuse through the entire cell to be collected by the p-n junction at the rear surface. High quality n-type FZ-Si is used as substrate. Furthermore, excellent front surface passivation is provided by thermal SiO<sub>2</sub> together with an n-doped front surface field. As an important high efficiency feature small-area contact holes are opened in the back passivation SiO<sub>2</sub> film and both p<sup>+</sup> emitter and n<sup>+</sup> base are contacted by metal fingers. With the novel low-cost production process Sun Power has achieved 21.1% efficiency on 149 cm<sup>2</sup> FZ silicon [15]. Construction of a cell factory in Manila in the Philippines with an annual capacity of 25 MW is underway [5].

### 3.2 The HIT Solar Cell

In 1997 a solar module was commercialized by Sanyo Electric, whose cells are based on a combination of monocrystalline and amorphous silicon [16]. Of prime importance are (i) the excellent surface passivation properties of intrinsic hydrogenated amorphous silicon (a-Si:H) and the use of n-type Cz-Si, which, as already discussed above, in addition to its high minority carrier lifetime, does not degrade under illumination despite oxygen is present. In Fig. 5 a schematic diagram of the HIT (Heterojunction with Intrinsic Thin-layer) solar cell is shown.



**Fig. 4.** Schematic diagram of the Point-Contact Solar Cell (taken from [15])

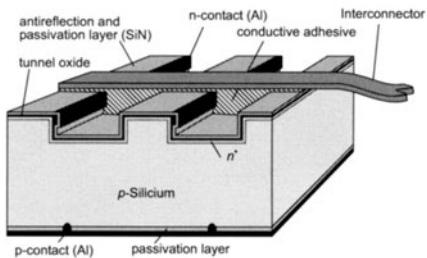


**Fig. 5.** Structure of the HIT solar cell (taken from [16])

The cell is composed of a textured n-type Cz-Si wafer sandwiched between p/i a-Si:H films on the illuminated side and i/n a-Si:H films as a back surface field (BSF) structure on the rear side. Transparent conductive oxide (TCO) layers are deposited on both sides followed by a silver grid for current collection.  $V_{oc}$  values exceeding 710 mV are due to the high quality of the a-Si / c-Si heterojunction including interface passivation. Low process temperatures of about 200°C and a reduced temperature coefficient of the cell are further advantages. Bifacial operation is possible. Limitations in the blue response of the cell may be given by the inherent absorption in the TCO layer and in the amorphous Si emitter layer. A record laboratory cell efficiency of 21.3% could be achieved on 100 cm<sup>2</sup> FZ using photolithography. In April 2003 a 200 W module was commercialized with 19.5% efficient cells resulting in a record module conversion efficiency of 17% [17].

### 3.3 The “Buried Contact” Solar Cell

Another candidate for high-efficiency commercial cells is the buried contact solar cell, developed already two decades ago and successfully transferred to industrial production in the early nineties [3,4]. A novel metalization scheme is applied, whereby laser grooves define the location and cross-sectional shape of the top surface metal conductors. The advantages are low grid shadowing by the fine top contact grid (20-30 μm wide), high fill factor due to low resistive losses in the metalization and low contact resistance. The front and rear contacts are formed by self-aligning electroless plating (Ni/Cu/Ag). The contact region is heavily phosphorous doped, whereas the rest of the surface is covered with a lowly doped emitter passivated with the silicon nitride antireflection coating. BP Solar is producing the “Saturn” solar cell based on the “buried contact” technology. Whereas up to now in the production line average efficiencies around 16% were achieved, only recently values of 18.3% on (12.5×12.5) cm<sup>2</sup> Cz-Si wafers could be reported [18]. Mass production of this new generation “Saturn” cells is planned for the year 2005.



**Fig. 6.** Schematic of the standard OECO cell structure. Busbar formation and tabbing are simultaneously performed using conductive adhesives

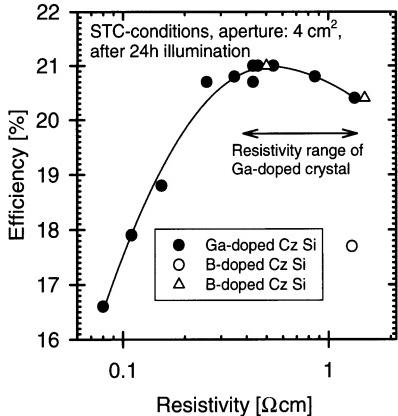
### 3.4 The Standard OECO Solar Cell

Recently, a new technology based on the OECO (Obliquely Evaporated COnact) process was entering the high efficiency area [19,20]. A schematic representation of the Standard OECO MIS-n<sup>+</sup>p solar cell is shown in Fig. 6. The main feature of the cell is the corrugated surface consisting of a set of parallel rectangular grooves. A shallow diffused n<sup>+</sup>-emitter is located at the silicon surface. The metal for the front grid fingers is obliquely evaporated upon an ultrathin low-temperature tunnel oxide on the upper part of the vertical flanks forming metal-insulator-silicon (MIS) contacts. They are an alternative to the sophisticated selective emitter contacts of the aforementioned cells [21]. For the oblique metal evaporation in a vacuum the self shading effect of the ridges is utilized, so that no masking or photolithography is required and high throughput is achieved since the wafers can be arranged closely spaced on a rotating cylinder [20].

On steep flanks high metal cross sections can easily be achieved without increasing the shadowing losses. The front surface of the cell is covered with low temperature (330°C) deposited PECVD silicon nitride acting both as passivation and antireflection coating. Whereas the front surface is randomly textured with pyramids, the vertical flanks can remain flat in order to keep the contact area low. This is simply accomplished by proper orientation of the grooves so that preferential etching does not occur at the flanks (e.g. for a (100) surface, the flank orientation should be of the (110) type).

by laser opening of the passivation layer and subsequent deposition of a continuous Al layer. The detailed processing sequence is described elsewhere together with the novel production equipment [19,20,22]. A new cell interconnection technology based on conductive adhesives has been developed with several advantages compared to soldering: i) low process temperature (< 200°C), ii) low mechanical stress and iii) lead-free.

In Table 1 the efficiencies after 24 h illumination of OECO MIS-n<sup>+</sup>p solar cells fabricated on different silicon materials are summarized. As an outstanding result, for (10×10) cm<sup>2</sup> OECO solar cells record efficiencies of 20.1% were achieved using Ga-doped Cz-Si. In Fig. 7 efficiencies of OECO solar cells (4 cm<sup>2</sup>) are presented for the resistivity range 0.08 Ωcm to 1.34 Ωcm of Ga-doped Cz silicon, some values for B-doped FZ-Si and Cz-Si are included for comparison. As can be seen, peak efficiencies of over 21% could be



**Fig. 7.** Efficiencies of OECO solar cells as a function of the substrate resistivity for Cz and FZ-Si with B and Ga dopants

obtained on 0.4  $\Omega$ cm Ga-doped material, which are comparable to those of Boron-doped FZ-Si, in particular no light induced degradation is observed. The result clearly indicates that the inherent axial resistivity variation of Ga-doped Cz-Si crystals is well acceptable for the fabrication of high efficiency solar cells [9]. In contrast, much lower efficiencies are resulting for B-doped solar grade Cz-Si due to light induced degradation.

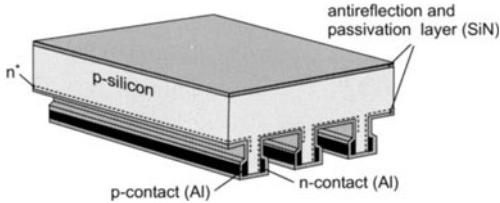
**Table 1.** One sun efficiencies (after 24 h illumination AM 1.5 G, 100 mW/cm<sup>2</sup>, 25°C) of 4 cm<sup>2</sup> and 100 cm<sup>2</sup> standard OECO MIS-n<sup>+</sup>p solar cells using different silicon materials (\*confirmed by FhG-ISE)

	FZ-Si		Cz-Si(Ga)		Cz-Si(B)	
Area (cm <sup>2</sup> )	4	100	4	100	4	100
Efficiency (%)	21.2*	20.0*	21.0	20.1*	18.3	17.9

### 3.5 Back Contacted Bifacial “BACK OECO” Solar Cell

With the second solar cell based on the OECO principle developed at ISFH both contacts are placed on the rear side, thus showing the same advantages which were already mentioned with respect to the back contacted Sun Power Cell. As a further very important characteristic, this cell is able to also efficiently utilize light impinging on the rear side, so that the output power can be significantly increased. The structure of the novel back contact device, designated by us as BACK OECO solar cell, is schematically shown in Fig. 8.

The characteristic features include: (i) A well passivated front surface by PECVD Si nitride, optionally with a floating junction (ii) A corrugated back surface with high quality MIS n-contact lines on one flank and p-contact lines on the opposite flank of the elevations. (iii) An n<sup>+</sup>-emitter at the back surface



**Fig. 8.** Structure of the BACK OECO solar cell. Both contacts are on the rear side. Also the light reflected onto the rear side can be utilized

except in the vicinity of the p-contacts, where a local back surface field can be introduced. (iv) The whole back surface is covered by a low temperature passivation layer (PECVD SiN). The high efficiency features apart from the metalization-free front side as well as the low cost production features are similar to those of the standard OECO cell and are outlined elsewhere together with the detailed processing sequence using only industrially feasible self aligning steps [23,24]. Due to contact metalization by oblique vacuum evaporation the separation of the n- and p-contact lines in order to avoid shunting is ingeniously simple and reliably accomplished without any masks or alignment. For the BACK OECO solar cell passivation of the front side by plasma silicon nitride plays a crucial role. This thin film material was investigated in the author's lab for many years and record low surface recombination velocities ( $< 4 \text{ cm/s}$ ) could be achieved [25,26].

As to the preliminary results,  $(2 \times 2) \text{ cm}^2$  cells were processed using  $0.5 \Omega\text{cm}$  FZ silicon wafers. By far not optimized an almost symmetrical behavior with efficiencies of 19% and 17.4% could be achieved for front and rear side illumination, respectively. Simulations reveal that efficiencies of 22% are attainable for high quality FZ-Si or Ga-doped Cz-Si with a diffusion length of about  $800 \mu\text{m}$  and a thickness between  $200 \mu\text{m}$  and  $300 \mu\text{m}$  [24]. With thinner wafers efficiencies above 20% can be obtained for lower quality material such as boron doped Cz-Si. As a further advantage of the OECO technology, the optimum base thickness can automatically be adjusted by the grooving process even below  $50 \mu\text{m}$  without wafer breakage since a supporting grid is provided by the groove ridges.

Utilizing the bifacial sensitivity of the BACK OECO solar cells in the novel multifunctional bifacial module arrangements recently introduced by us, additional power gains of more than 50% are resulting compared to monofacial use [28]. A power output per cell is expected to be equivalent to that of an at least 30% efficient monofacial cell of the same size. Thus bifacial cells properly installed at low cost are promising candidates to make PV significantly more economic.

## 4 Conclusion

The crystalline silicon solar cells presented in this paper have the potential to reach efficiencies even beyond 20% in an industrial environment, thus

opposing the general opinion that high efficiencies can only be realized using complex processing. Utilizing the bifacial option of the BACK OECO solar cells, additional power gains of more than 50% can be achieved, resulting in further significant reduction of the cost per watt ratio. To our opinion, in the future crystalline silicon and thin film technologies will share the PV market. The efficiency region around 20%, required for power applications with limited area, will be the domain of crystalline Si solar cells, whereas the thin film cells will occupy the efficiency range up to 14%. Due to its almost unlimited availability, silicon will most probably dominate the PV market also in the long term.

## References

1. S.R. Wenham, M.A. Green, M.E. Watt, *Applied Photovoltaics*, Centre for Photovoltaic Devices and Systems, UNSW, Sydney.
2. J. Zhao, A. Wang, M.A. Green, Prog. Photovolt.: Res. Appl., **7**, 411 (1999).
3. M.A. Green, Prog. Photovolt. Res. Appl. **8**, 127 (2000).
4. M.A. Green, *Silicon Solar Cells: Advanced Principles and Practice*, Bridge Printery, Sydney, (1995).
5. J. Bernreuter, Photon International, May 2003.
6. J. Knobloch, et al., Proc. 13th Europ. Photovolt. Solar Energy Conf. (Stephens, Bedford), 9 (1995).
7. J. Schmidt, A.G. Aberle, R. Hezel, Proc. 26th IEEE Photov. Spec. Conf., Anaheim, CA, IEEE, New York, 13 (1997).
8. J. Schmidt, K. Bothe, R. Hezel, Proc. 3rd World Conference on Photov. Energy Conversion, WCPEC-3 Osaka, in press (2003).
9. A. Metz, T. Abe, R. Hezel, Proc. 16th Europ. Photov. Solar Energy Conference, Glasgow, UK, (James and James, London), 1189 (2000).
10. S.W. Glunz, S. Rein, J.Y. Lee, W. Warta, J. Appl. Phys. **90**, 2397 (2001).
11. J. Schmidt, Solid State Phenomena **95-96**, 187 (2004).
12. K.A. Münzer, et al., Proc. 2nd World Conf. on Photovoltaic Energy Conversion, Vienna, 1214 (1998).
13. J. Vedde, T. Clausen, L. Jensen, Proc. 3rd World Conference on Photov. Energy Conversion, WCPEC-3, Osaka, in press (2003).
14. P.J. Verlinden, et al., Proc. 12th European Photov. Solar Energy Conference, 1304 (1994).
15. K.R. McIntosh, et al., Proc. 3rd World Conf. on Photovoltaic Energy Conversion, Osaka, in press (2003).
16. H. Sakata, et al., Proc. 3rd World Conf. on Photov. Energy Conversion, WCPEC-3, Osaka, in press (2003).
17. T. Sawada, et al., Proc. 1st World Conf. on Photov. Energy Conversion, WCPEC-1, Hawaii, 1219 (1994).
18. T.M. Bruton, et al., Proc. 3rd World Conference on Photov. Energy Conversion, WCPEC-3 Osaka, in press (2003).
19. R. Hezel, Ch. Schmiga, A. Metz, Proc. 28th IEEE Photov. Spec. Conf. Anchorage, 184 (1999).
20. R. Hezel, Solar Energy Materials and Solar Cells **74**, 25 (2002).

21. R. Hezel, Recent Progress in MIS Solar Cells, *Progr. Photovolt. Res. Appl.* **5**, 109 (1997).
22. A. Metz, R. Hezel, *Solar Energy Mater. Sol. Cells* **65**, 325 (2001).
23. J.W. Mueller, A. Merkle, R. Hezel, Proc. PV in Europe, Conference and Exhibition, Rome, 48 (2002).
24. R. Hezel and W. Hoffmann, Proc. 3rd World Conf. on Photov. Energy Conversion, WCPEC-3, Osaka, in press (2003).
25. R. Hezel, K. Jaeger, *J. Electrochem. Soc.* **136**, 518 (1989).
26. T. Lauinger, et al., *J. Vac. Sci Technol. A* **16**, 530 (1998).
27. R. Hezel, Proc. 29th IEEE Photov. Spec. Conference, New Orleans, 114 (2002).
28. R. Hezel, *Prog. Photovolt.: Res. Appl.* **11**, 549 (2003).

# Second and Third Generation Photovoltaics – Dreams and Reality

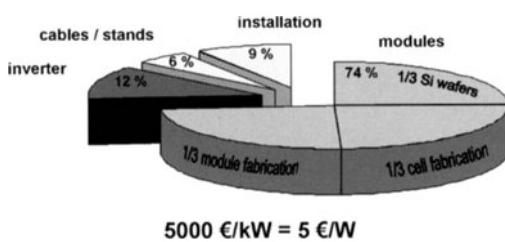
Jürgen H. Werner

Institut für Physikalische Elektronik, Universität Stuttgart  
Pfaffenwaldring 47, 70569 Stuttgart, Germany

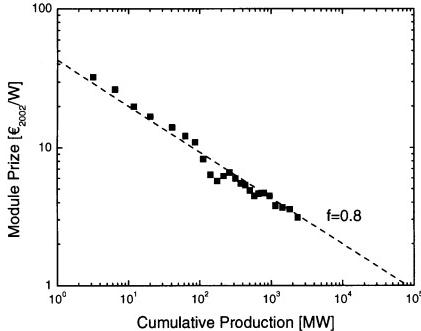
**Abstract.** The photovoltaic market of 750 MW/year is dominated by Si wafers and grows with an annual rate of around 30 %. At the beginning of 2004, about 350 MW of photovoltaic power is installed in Germany. Based on an annual growth of 20 to 30 %, between 8 and 30 GW will be installed in 2020. The production cost of the crystalline Si modules halve about every ten years. At present, solar electricity costs 0.6 EURO/kWh in Germany. Provided the mass production continues, this number may drop to about 0.2 EURO/kWh in 2020. Classic materials of so-called second generation photovoltaics such as amorphous Si, CdTe, Cu(In,Ga)Se<sub>2</sub> have either physical, ecological or technological disadvantages. Only crystalline Si itself, in thinner form (either as thinner wafers or as a thin film) will probably be able to add to or to replace the technology of about 200 to 300  $\mu\text{m}$  thick Si wafers. New physical concepts in photovoltaics, so-called third generation photovoltaics, could lead to photovoltaic conversion efficiencies of above 30 %. However, also in this case, concepts that make use of crystalline silicon seem to be the most promising.

## 1 Introduction

Figure 1 shows the end-user price for a typical photovoltaic (PV) roof top system as it will be installed in the year 2004 in Germany. Without taxes, the total costs (including installation) amount to about 5000 EURO/W. Two thirds of this sum go into the modules which are sold to the end user for about 3.5 EURO/W. Depending on the size of the photovoltaic system, the insolation, the interest rate and the annuity of the credit for the photovoltaic system, the generation cost of photovoltaic electricity in Germany amounts to about 0.60 EURO/kWh if the roof top system is financed, depreciated and used over a time of twenty years [1,2].



**Fig. 1.** Cost for a 2 kW roof top system to be installed in Germany in 2004. Modules make up the largest fraction in the cost and are mainly responsible for the cost of photovoltaic electricity



**Fig. 2.** Learning curve of crystalline silicon modules. At present, about 2 GW of PV modules are installed world-wide. The total production cost range around 3 EURO/W for crystalline Si and decreases with a cost reduction factor  $f = 0.8$ , meaning a cost reduction of 20 % for every doubling of the cumulative production. Extrapolating into the future yields a cost of 1 EURO/W for 70 GW production

Most of the roof top systems which are installed world-wide contain crystalline silicon either in the form of single-crystalline or polycrystalline wafers with a typical thickness of 300  $\mu\text{m}$ . Photovoltaics based on wafered crystalline silicon has been termed as *first generation* PV. *Second generation* PV does not make use of 200 to 300  $\mu\text{m}$  thick crystalline silicon wafers but of 1 to 20  $\mu\text{m}$  thin films, either of *amorphous* Si (a-Si), polycrystalline CdTe or one of the alloys of Cu(In,Ga)(S,Se)<sub>2</sub>. Recently, first modules based on the "new" thin-film material crystalline silicon entered the market [3]. *Third generation* PV aims at overcoming the classic Shockley/Queisser efficiency limit [4], which ranges at a maximum theoretical efficiency of 30 %.

This contribution discusses the status of first, second, and third generation PV. At present, second generation PV cannot meet the expectations for a faster cost reduction for PV electricity than with first generation, silicon wafer-based PV. Only crystalline silicon itself in the form of thin films might be able, in the future, to compete with or to replace silicon wafers.

Third generation PV is of high interest for basic research. Apart from the classic concepts of either tandem cells or concentrating and tracking the sunlight, most of the concepts are still speculative. Thus, Si wafer-based PV will dominate the market for a much longer time than it was assumed about ten years ago.

## 2 Learning Curve and Module Market

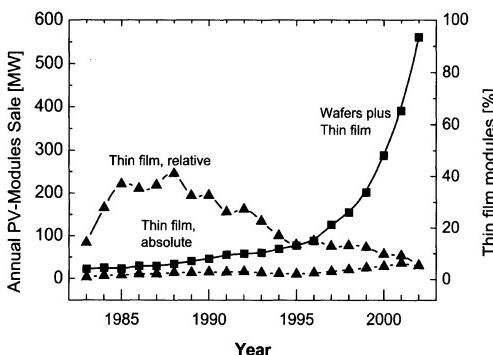
### 2.1 Learning Curve of Crystalline Silicon Modules

The retail prizes for photovoltaic modules in Fig. 2 demonstrate that the cost degradation of classic crystalline silicon PV modules follows a power law [5]. The cost reduction of PV modules differs in no way from any other industrial good: The more one produces, the lower the cost per piece. The higher the cumulative production, i.e. the sum of all goods produced before, the lower the cost. For the production cost  $C_n$  of the  $n$ -th series of the modules it holds the empirical law  $C_n = AP_a^{-\beta}$  Here  $P_a$  stands for the cumulative production,

whereas  $A$  is the cost at the beginning and  $\beta$  the cost reduction rate. When the cumulative production doubles from  $P_a$  to  $P_b = 2 P_a$ , then it holds for the ratio  $C_b/C_a$  of the respective production cost that  $\frac{C_b}{C_a} = (P_b/P_a)^{-\beta} = 2^{-\beta} = f$ , where the quantity  $f$  is the learning factor. For the curve in Fig. 2, it holds  $f = 0.8$ . For a doubling of the cumulative production, the cost dropped in the past from  $C_a$  to  $C_b = fC_a = 0.8 C_a$ , this means by  $f' = 1 - f = 20\%$  for every doubling.

## 2.2 Market Growth

Figure 3 displays the exponential growth of the module sales [6]. Within the last ten years, the module market grew with an annual rate of about 30 %. This growth was almost exclusively carried by modules from wafered Si; the share of sold second generation thin film modules dropped from 40 % in 1988 to 5 % in 2002.



**Fig. 3.** The PV module market grows exponentially with an annual increase of 30 %. Thin-film modules (a-Si, CdTe, CIGS) make up only about 30 MW/year

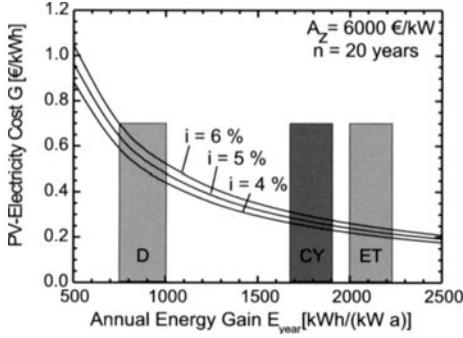
## 3 Electricity Cost

### 3.1 Dependence on Insolation

Figure 4 presents the generation cost  $G$  of photovoltaic electricity, which depends on the total cost  $A_z$  per kW of the PV system over the time of its operation (including all repair and service) over  $n$  years, the interest rate  $i$  for the loan that is taken to finance the system, and the annuity  $K = i/[1 - (1 + i)^{-n}]$  according to

$$G = \frac{A_z K}{E_{\text{year}}} \quad (1)$$

The quantity  $E_{\text{year}}$  stands for the annual energy collection. For the South of Germany, it holds  $E_{\text{year}} \approx 1000 \text{ kWh}/(\text{kW a})$ , whereas for Cyprus  $E_{\text{year}} \approx$



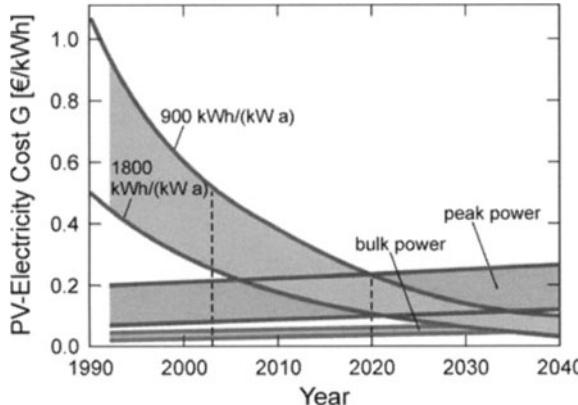
**Fig. 4.** The generation cost of photovoltaic electricity from Eq. (1) depends not only on insolation, but also on the interest rate  $i$  of the loan that finances the PV system of cost  $A_z$  over the depreciation (and operation) time of  $n = 20$  years

1700 kWh/(kW a), and Egypt  $E_{\text{year}} \approx 2100$  kWh/(kW a). In case of PV systems, the cost for maintenance and repair are low (about 50 EURO/kW a) during the operation; the largest share of the cost is made up by the PV system itself. As a consequence (see Fig. 1), modules make up the largest fraction of  $A_z$ . Reduction of module fabrication cost (by following or improving the learning curve in Fig. 2) is of utmost importance. According to Fig. 4, at present the cost for PV electricity ranges at around 0.60 EURO/kWh in Germany and at 0.30 EURO/kWh in the Southern part of Europe. System prizes around 6000 EURO/kW (5000 EURO/kW for the system and 50 EURO/year fixed cost over 20 years) are typical.

### 3.2 Cost Degradation

Predictions on the time dependence of PV electricity cost require assumptions about the prize reductions of PV modules. Prizes will reduce when the market continues to grow as presented in Fig. 3, and if it is allowed to extrapolate the empirical learning curve in Fig. 2 towards larger values of the cumulative power. The production processes have to become cheaper by automation as well as by new results from research and development. The curves in Fig. 5, taken from Hoffmann [1], assume a market growth of 22.5 % until the year 2020 and start with a PV electricity cost of 0.60 EURO/kWh in the year 2000. According to the upper curve in Fig. 5, for the illumination conditions in the Northern part of Germany with an annual energy gain of  $E_{\text{year}} \approx 900$  kWh/(kW a), PV-electricity in Germany will have a generation cost of 0.20 EURO/kWh.

Roughly speaking, around 2020, for German illumination conditions and for the end-user, PV electricity generated on his house (and consumed there) has about the same cost as is the present prize for electricity bought from the utility. In the Southern part of Europe, the generation cost of PV-electricity is about half of that in Germany. In any case, between 2020 and 2030, electricity generated by PV will – even for a utility – reach the market for peak power. Assuming a 20 % (30 %) annual growth for the installed PV power in



**Fig. 5.** Cost decrease of photovoltaically generated electricity. For a market growth of 22.5 %, Hoffmann predicted about 0.20 EURO/kWh generation cost to be reached around the year 2020 for German insolation (here taken as 900 kWh/(kW a)). At this cost level, photovoltaic electricity is able to penetrate the market for peak power. In Southern Europe (1800 kWh/(kW a)) the generation cost is lower

Germany, yields 1.2 GW (2.2 GW) PV power for the year 2010 and 7.8 GW (30 GW) for the year 2020.

### 3.3 Areas and Production Rate of PV Modules

In Germany, public power plants with a total peak power of around  $P_{\text{total}} = 105.8 \text{ GW}$  [7] are installed; the total German annual electricity consumption is  $E_{\text{total}} = 534 \text{ TWh/a}$ . On the average, the public power plants operate over 5000 h of the 8760 h per year at full power. In contrast, solar modules operate only about 1000 h per year at full power: At optimum geometry, the sun delivers an input of  $1000 \text{ W/m}^2$ , which together with a system efficiency of 10 % yields an areal output of  $100 \text{ W/m}^2$ . The solar input is about  $1000 \text{ kWh/m}^2\text{a}$ , which is the same as if the sun shone for 1000 h at full power. The electricity output of the whole PV system is about  $E_{\text{year}} = 1000 \text{ kWh/(kW a)}$ .

Without storage systems, photovoltaic electricity is only able to contribute a part of the electricity supply. However, it should be possible to supply at least about 10 % of the annual German electricity consumption, i.e. 50 TWh/a. With an energy collection  $E_{\text{year}} = 1000 \text{ kWh/(kW a)}$ , the required PV peak power would be 50 GW. With intelligent control technologies, on the long run, it should be possible to master such a power. Even at present, the consumed power in Germany varies between 55 and 80 GW, i.e. over more than 25 GW. The required 50 GW of installed PV modules need about  $500 \text{ km}^2$  of area, which is readily available; the total roof area in Germany is  $3000 \text{ km}^2$ . Thus, it seems possible to supply a remarkable part of German electricity with PV, *but not with every PV material*.

At present the guaranteed lifetime of PV modules is 20 years (and 40 years seem possible). For the assumed PV power of 50 GW (which could be the case around 2030), every year 2.5 GW PV modules would have to be replaced [8]. For such a large number of modules for some technologies, e.g. Cu(In,Ga)Se<sub>2</sub>, *supply and availability* of raw materials might be critical. For silicon-based technologies there is no problem.

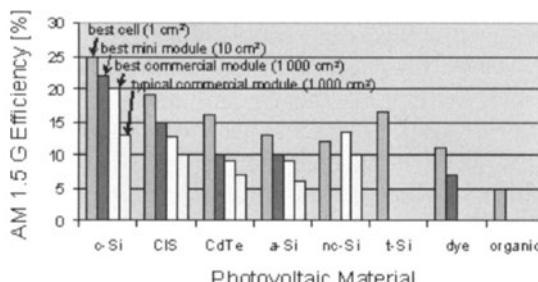
## 4 Second Generation Hopes: Thin-Film Materials

Second generation PV lives from the idea of making the production of modules cheaper by consuming less semiconducting material. In the modules, highly absorbing semiconductors are used. They require only thicknesses of several  $\mu\text{m}$  to absorb all of the sunlight, and one uses integrated processing in the module fabrication. According to Fig. 1, the fabrication cost of Si wafers and cells alone makes up about half of the cost of a complete PV system. Second generation modules avoid Si wafer fabrication.

At present, four kinds of thin-film modules are commercially available: amorphous Si (a-Si), CdTe, Cu(In,Ga)Se<sub>2</sub> (CIGS) and (nano)crystalline Si. Figure 6 compares the status of thin-film cells and modules with Si wafer technology. Apart from Si-, CdTe- and CIGS-based technologies, I include also data for injection-type and organic cells.

Table 1 presents a cost comparison between different technologies: Czochralski-grown single-crystalline Si, polycrystalline Si wafers from cast blocks, a-Si, CdTe, and Cu(In,Ga)Se<sub>2</sub>. Frantzis [9] et al. listed data for factories with assumed equal production rates of 10 MW/a and 100 MW/a. At present, factories for modules from wafered single-crystalline or polycrystalline Si are going towards the 100 MW/a production. The largest factories for a-Si and CdTe (only one) are just at the 10 MW/a level; in contrast, CIGS is at the 1 MW/a level. The data of Frantzis were also discussed by Schott [10].

According to Table 1, thin-film technologies suffer not only from a low efficiency, but, in particular, from a low yield in the production of the modules.



**Fig. 6.** AM 1.5 G efficiencies for different second generation thin-film materials. In addition, data for dye-sensitized as well as for organic cells are shown

**Table 1.** Degradation of direct manufacturing cost  $C$  for thin film modules from a-Si, CdTe, and Cu(In,Ga)Se<sub>2</sub> in comparison to Czochralski (Cz)-grown single crystalline Si wafer, and polycrystalline cast Si wafer based technology. Data in columns two and three are taken from Frantzis et. al. [9] The learning factor  $\gamma$  in column 4 follows from Eq. (2). Columns five and six show the assumed parameters [9] which go hand-in-hand with the cost degradation by increasing the factory sizes from  $S = 10 \text{ MW/a}$  to  $S = 100 \text{ MW/a}$

Photovoltaic material	$C$ [\$/W], $S = 10 \text{ MW/a}$	$C$ [\$/W], $S = 100 \text{ MW/a}$	learning factor $\gamma$	module efficiency $\eta$ [%]	manufacturing yield [%]
Cz-Si	2.45	1.45	-0.228	15 → 18	85 → 95
Cast Poly-Si	2.12	1.15	-0.266	14 → 17	85 → 95
a-Si	2.30	1.40	-0.285	6 → 8	75 → 90
CdTe	2.30	0.95	-0.384	8 → 12	65 → 90
CIGS	2.25	1.00	-0.352	9 → 13	65 → 90

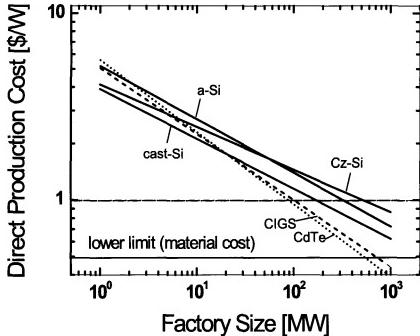
Defective Si wafers are separated out either in the wafering process itself or on the cell level. In case of a thin-film module, a single small-area defect can result in a failure of the whole module of square meter size.

The data in Table 1 elucidate that all thin-film technologies, at present, have *higher* production cost than Si wafer-based technologies on the 10 MW/a level, and that thin-film technologies might only have a chance for *lower* cost on the 100 MW/a level.

Figure 7 is based on the data of Table 1, and rigorously uses the data. Here I assume that the fabrication cost  $C$  for a module depends on the size  $S$  of a factory in a similar way as the fabrication cost depends on the cumulated production in Fig. 2, i.e. according to a power law,

$$C = BS^{-\gamma}. \quad (2)$$

Here  $\gamma$  is the cost reduction rate which is different for every technology and  $B$  is a constant (also depending on technology). According to Fig. 7, CIGS and CdTe start at relatively high cost but their cost decreases faster. Thus, at first sight, Fig. 7 suggests that thin-film materials listed there might be able to beat not only Cz-wafer Si technology, but also cast Si wafers. The intersection of the lines for CIGS and cast poly-Si lies at a 20 MW production line. However, the question arises, when (and if ever) the three thin-film materials a-Si, CdTe, Cu(In,Ga)Se<sub>2</sub> will reach such large production rates! Today, no 10 MW production line of CIGS exists, whereas first productions of 100 MW/a of c-Si modules are in sight. Thus, c-Si wafer technology drives down along the learning curve much faster than it is possible to build up the production for thin-film modules. A 200 MW/a production for c-Si wafer cells (at the 1.0 \$/W cost level in Fig. 7) will probably be available much earlier than a 20 MW/a production for CIGS (1.8 \$/W in Fig. 7), which is the only material that could compete with the efficiency of c-Si. The horizontal line



**Fig. 7.** For a factory with larger size (and progress in yield and efficiency) the cost per module are lower. Curves in this figure base on the data from Tab. 1. Only for 10 MW/a and 100 MW/a data are available; all other data follow by extrapolating on the basis of (2)

at 0.5 EURO/W in Fig. 7 considers that lowering the cost is limited at by material cost at this level [11].

#### 4.1 Amorphous Si (a-Si)

The world-best small-area ( $1 \text{ cm}^2$ ) a-Si cell containing three different amorphous layers shows an initial efficiency of 15.2 % under AM 1.5 G illumination and degrades to a stable efficiency of 13 % after 1000 h of illumination. Deng and Schiff gave a recent review on a-Si technology [13].

According to Schade and Maurus, commercial large area modules are about 5 to 6 % efficient in the stable state [12]. Despite about 20 years of research, the technology of amorphous Si has not overcome the inherent efficiency limitations of a-Si which originate in defect states within the forbidden gap of a-Si. Hydrogen is able to passivate many of these defects, but not all [13].

According to Table 1, the production cost for a-Si modules are higher than those for crystalline Si. Therefore, a-Si finds only a market, where it has an intrinsic advantage. Niche markets for a-Si are

*Application in flexible modules.* The deposition temperatures of a-Si range around 250 °C. Thus it is possible to deposit on plastics or on steel.

*Applications under low illumination levels.* The high band gap  $E_g = 1.7 \text{ eV}$  yields low saturation current densities and high shunt resistances. Therefore, the efficiency of a-Si cells for indoor illumination is better than for mediocre crystalline cells. This leaves a market for a-Si in wrist watches, pocket calculators, etc. Gemmer modeled the behavior of a-Si (as well as Cu(In,Ga)Se<sub>2</sub> and c-Si) under these conditions [14].

*Applications at high temperature.* The large band gap of a-Si is reflected in a relatively small degradation of the efficiency (0.3 %/K), when compared to c-Si (0.5%/K) for increased temperatures.

#### 4.2 CdTe

A CdTe-module of  $1 \text{ m}^2$  with a power of 100 W contains about 10 g of the heavy metal Cd, best cells have an efficiency of 16.5 % [15]. Commercial large-

area modules produced by the American company First Solar (2 MW/a production capacity) or the German ANTEC (10 MW/a production capacity) range from 7 to 8 %. BP Solar (10 MW/a capacity) in the USA as well as Matsushita Battery (10 MW/a capacity) in Japan gave up their production in 2002.

Reasoning for CdTe is often based on statements like "CdTe is the ideal solar cell material" [16], going back to a calculation of Loferski of 1956 [17], who – on the basis of the material quality of that time - calculated a maximum possible solar cell efficiency of around 25 % with an optimum band gap of 1.5 eV, close to the band gap of CdTe. For silicon he predicted an upper limit of 22 %. Not only recent experimental data (Si reached almost 25 % efficiency [18]), but also the classic calculation of Shockley and Queisser of 1961 [4] demonstrated that the reasoning of Loferski [17] cannot be taken to select a material on the basis of the band gap value only. Instead one has to require perfect electronic quality [4]. Apart from the unavoidable radiative recombination, all other recombination processes have to be suppressed. With 16 % efficiency, CdTe is far away from that requirement.

The CdTe in solar modules is polycrystalline. Similarly to other polycrystalline materials, recombination at grain boundaries is probably the reason for the relatively low electronic quality. It is hard to imagine that polycrystalline CdTe cells will ever reach the 20 % mark. Werner, Mattheis, and Rau [19] as well as Rau and Werner [20] demonstrated recently, that spatial potential fluctuations within the bulk as well as at grain boundaries are responsible for the lower efficiency in polycrystalline materials, even if the carriers were subjected to radiative recombination only.

Compared to crystalline silicon, CdTe benefits from two properties.

*Simple fabrication technology.* The manufacturing temperatures lie below 600°C and consume only a small amount of energy. In addition, the deposition processes are fast. Despite the simple processing, the quality of the material is reasonable [16].

*Large band gap value.* Similarly to a-Si, the large band gap  $E_g = 1.5$  eV promises lower temperature coefficients for the efficiency and higher annual yield collection than for c-Si. Concerning the second point, the literature is contradictory and states data between 700 kW/(kW a) to 1100 kWh/(kW a) for German illumination conditions, compared to a value of 1000 kWh/(kW a) for c-Si modules. More detailed measurements are necessary.

Apart from grain boundaries, the main obstacle in achieving higher module efficiencies lies in problems with the fabrication of low-resistive back contacts resulting in low fill factors and module efficiencies below 10 %.

### 4.3 Cu(In,Ga)Se<sub>2</sub> (CIGS)

With more than 19 % efficiency under standard illumination conditions and a cell area of 0.5 cm<sup>2</sup>, Cu(In,Ga)Se<sub>2</sub> (CIGS) showed the best performance of all thin-film solar cell materials [21,22,23]. Best CIGS films are prepared

by simultaneously evaporating the elements Cu, In, Ga, Se at  $T = 550\text{ }^\circ\text{C}$  on glass substrates. Würth Solar (1.2 MW/a capacity) and Shell Solar (using a slightly different process) manufacture modules of  $60 \times 120\text{ cm}^2$  size with up to 12 % efficiency (best modules) and typical module efficiencies of 10 %.

It is possible to vary the band gap in the alloy system  $\text{Cu}(\text{In}_{1-x}\text{Ga}_x)\text{Se}_2$  over a relatively wide range. Unfortunately, films with a high or low Ga content show low performance, the best cells need a Ga content  $x = 0.23$  with a band gap of  $E_g = 1.15\text{ eV}$ , which is almost the same band gap as the one for crystalline silicon. Thus, at present it does not seem possible to fabricate highly efficient tandem cell structures on the basis of this material system. The following physical, ecological and economic questions are open.

*Efficiency limitations by fundamental material properties.* Films from CIGS are polycrystalline with grain sizes below  $1\text{ }\mu\text{m}$ . The small grains, the high compensation ratio as well as compositional variations lead not only to an inherent structural, but also to an inherent *electronic* inhomogeneity. Inhomogeneities on all length scales might limit the efficiencies [19,20]. Even small-area cells will hardly master the 20 % efficiency mark [19]. *Instabilities under illumination.* CIGS-modules improve under operation. Here many questions are open, including the statement of module specification.

*Thin film protection.* CIGS modules degrade in water vapor. Therefore, the sealing of the modules is critical.

*Ecological questions.* The element Se in CIGS has a relatively low biological activity but many of its compounds (in particular the gas  $\text{H}_2\text{Se}$  which is used in some production processes) are highly active and toxic. Up to now, all modules need a CdS buffer on top of the CIGS before the ZnO front contact is deposited. Research tries to replace this buffer.

*Abundance of In.* According to Anderson [26], for a 1 GW/a production of CIGS, the present technology would need 20 % of the world-wide annual In production. In this case, In for CIGS module production would certainly compete with In consumption in other areas, for example, in the display market as well as with In for InGaN luminescence diodes. As a consequence, In prizes might increase. The relatively low and scattered abundance of In limits the total number of CIGS PV systems that could be installed world-wide: to 90 GW [26]. At present the world-wide electricity production is 14 000 TWh/a. Assuming 1000 h/a operation at full power, the 90 GW of possible CIGS modules could just produce 0.6 % of this mark.

#### 4.4 Crystalline Si (c-Si)

A thin-film technology based on crystalline Si could combine all the advantages of crystalline Si (material availability, non-toxicity, proven high efficiencies, possibility for using the whole knowledge of the microelectronic industry as well as from thin-film transistors etc.) with the advantages of a thin-film technique (large-area deposition, materials savings). However, compared to

CdTe and CIGS, it is much more difficult to prepare material of high electronic quality when one deposits Si in polycrystalline form on a substrate such as glass. As soon as the grains become smaller than about  $100\text{ }\mu\text{m}$ , the efficiency of solar cells degrades dramatically below 10 %. Research on crystalline silicon thin-film cells was reviewed by Bergmann [27], Werner [28,29], and Sopori [30]. It seems that there are only two ways to achieve good efficiencies.

*Nanocrystalline Si on glass.* Low-temperature deposition processes at  $400^\circ\text{C}$  and below allow one to deposit Si from gas with a high hydrogen content. For deposition rates of around  $0.5\text{ nm/s}$ , the deposited Si atoms have enough time (and mobility) to saturate their bonds with neighboring Si atoms or hydrogen. At least  $2\text{ }\mu\text{m}$  thick films are necessary to absorb enough light in crystalline Si. The low deposition rate thus leads to deposition times of more than an hour. Increasing the deposition rate results in lower material quality; the efficiencies degrade below 10 %. The Japanese company Kaneka uses such a technology [31]. The grain sizes for such kinds of technologies range between  $10\text{ nm}$  and  $1\text{ }\mu\text{m}$ . Werner [28] as well as Taretto, Rau, and Werner [32] analyzed many published data on polycrystalline cells and made a particular kind of electronically low-active [110] tilt boundaries responsible for the relatively good properties of this fine-grained silicon. Nevertheless, the electronic quality of low-temperature deposited nano- and microcrystalline silicon is inferior to the properties of CIGS. The highest efficiencies for small-area cells range at 12 %, compared to the more than 19 % of CIGS. Combined with top cells from amorphous Si, the Japanese company Kaneka fabricates modules with (non-stabilized) efficiencies of 11 % from nanocrystalline silicon [31].

*Single Crystal Si films or slivers.* There were many experiments to deposit Si at higher temperatures than  $400\text{ }^\circ\text{C}$  on a foreign substrate [27,28,30]. At such high temperatures, the films do not contain hydrogen and they grow much faster. Unfortunately, the electronic quality of the films is always low, even when the grain size is much larger (up to several  $\mu\text{m}$ ) than material from the low-temperature deposition. Obviously, the grain boundaries have inferior properties [28,29]. Therefore, one tries to increase the grain size further by melting the films either with electron beams or with lasers. Nevertheless, the highest efficiency was thus far limited to about 16 % for small cells [27,28]. In addition, such a heating requires a high-temperature substrate. In principle, one tries to come as close as possible to the properties of single-crystalline Si with such an approach. Consequently, one could start with single-crystalline Si right away. This approach seems also to be the only one which could yield similar efficiencies for thin-film Si solar cells as for CIGS solar cells. At present, two approaches are under investigation for single-crystalline Si films: i) transfer cells. In this case, one peels off thin films from the surface of a single-crystalline Si wafer and transfers them onto a foreign substrate. The world-best cells with more than 16 % efficiency are fabricated at the University of Stuttgart [33,34]. A group at the University of Erlangen works

**Table 2.** Maximum possible efficiencies in the limit of radiative recombination. The AM 0 spectrum of the cell is distributed over  $m$  cells. For maximum concentration  $C_{\text{conc}} = 46300$ , the cell sees the sun under a solid angle of  $2\pi$

Number $m$ of perfect semiconductors	$\eta_{\text{max}} [\%]$ , $C_{\text{conc}} = 1$	$\eta_{\text{max}} [\%]$ , $C_{\text{conc}} = 46300$
1	30	41
2	43	56
3	49	64
$\infty$	68	86

on a similar technique [35]. ii) In 2003, the National University of Canberra, Australia, reported a new technology, SLIVER-cells, which is now going into production. Here one prepares many small slivers from a 1 mm thick Si wafer and uses light trapping in order to consume as less silicon as possible [36]. The efficiency of first modules is above 12 % [37]. At least for single small-area cells, either the transfer approach or the SLIVER technology has the potential to yield cells with 20 % efficiency.

## 5 Third Generation Dreams: New Physics

Solar Cell efficiencies of above 30 % are possible if the spectrum of the sun is distributed over several different semiconductors with different band gaps. For radiative recombination being the only loss process, 68 % conversion efficiency is theoretically possible for a non-concentrated AM0 spectrum. Table 2 lists the maximum possible efficiencies  $\eta_{\text{max}}$  [38]. For maximum concentration  $C_{\text{conc}} = 46300$ , up to 86 % conversion efficiency are thermodynamically possible. Note, however, that this number implies the assumption that a very large number of perfect semiconductor materials (suffering only from radiative recombination) were available.

Research on efficiencies far above 30 % and the "third generation PV" started in 1994 after Werner, Kolodinski, and Queisser [39] as well as Werner, Brendel, and Queisser [40] demonstrated that, theoretically, efficiencies above 30 % would be possible even with a single semiconductor material, if the band structure could be optimized for impact ionization by high energetic photons from the solar spectrum. Since then, several effects have been proposed for highly efficient solar cells. The research of the last years has been reviewed by Green [41] as well as by Marti and Luque [42]. There are five main concepts for efficiency increases.

*Stacks of multiple cells, tandem cells.* The GaInP/GaAs/Ge cells of King et. al. [43] convert the AM 1.5 G sunlight with 33 % efficiency without any concentration. Unfortunately, the fabrication cost for these GaAs-based cells is too high for large-area terrestrial application. PV could benefit a lot from

tandem cells, if there was a top material for the "working horse" crystalline Si available. Unfortunately, at present no semiconductor of high electronic quality fits to the small lattice constant of Si.

*Concentrator cells.* For a concentration ratio of  $C_{conc} = 66$ , the cells of King et al. [43] yield an efficiency of 35.2%; an optical system tracks the sun. Thus, only radiation coming directly from the sun's disc is converted. In Germany, with respect to annual global incidence (annual irradiation onto a horizontal surface), the contribution of direct sunlight is about 40 % only.

*Utilization of low-energy photons.* Photons with energies  $h\nu < E_g$  are not absorbed in a semiconductor with a single band gap  $E_g$ . Already Gütterl and Queisser [45] investigated the possibility of efficiency gains via multi photon excitations via states within the band gap of silicon cells. At that time, these authors concluded that there would be no efficiency gain. Recently, Luque and Marti [46] as well as Marti, Cuandra and Luque [47] showed that under certain conditions (only radiative transitions) efficiency gains would be possible. At present there is no material in sight that fulfills the necessary physical requirements.

*Hot electron cells.* In a normal solar cell, the part of the photon energy which exceeds the band gap energy is waisted into heat. Werner, Kolodinski, and Queisser [39] demonstrated optically induced carrier multiplication for photon energies above the direct band gap (3.4 eV) of highly efficient crystalline Si solar cells. High-energy photons create hot electrons which, in turn, generate a second electron/hole pair via impact ionization. In their Si solar cells, up to 30 % of the photons with  $h\nu > 3.4$  eV gave rise to two electron/hole pair. Unfortunately, the solar spectrum contains only few photons with such large energies. The gain of the short circuit current density in Si cells is too low. It remains open if a semiconductor could be developed which shows efficient carrier multiplication in the regime of the visible, e.g. around 2 eV. Würfel argued that more efficient cells could be fabricated if the hot electrons could be extracted at the contacts before impact ionizing secondary electron/hole pairs [48]. In any case, it seems easier to develop Si-based tandem cells, instead of optimizing band structures for optimum impact ionization processes.

*Photon conversion.* Instead of using the polychromatic spectrum of the sun in either one or more semiconductors, it looks seducing to convert the sunlight and to adjust the spectrum to the semiconducting converter. Trupke, Green and Würfel discussed up- and down-conversion [49,50]. Unfortunately, at present, no really efficient systems for the optical conversion are available.

## 6 Conclusions

Compared to ten years ago, PV has become commercial business. With more than 700 MW of modules sales in 2003 and an estimated (world) averaged sales prize of 7 EURO/W, the annual turnover of PV industry is almost 5 billion EURO. Sales grow with an annual rate of 30 %; the cost of modules

halve about every ten years. Crystalline Si-based modules make up 95 % of the module market and rapidly run down the learning curve. First fabrication plants of 100 MW annual fabrication capacity are in sight. In contrast, classic second-generation photovoltaic materials such as a-Si, CdTe, and CIGS have only a small market share. Under such conditions, it is hard to imagine that either CdTe or CIGS could gain a large share. CdTe has severe ecological disadvantages, those for CIGS are not severe, but they exist. On the one hand, CIGS has advantages when compared to crystalline Si: The grain boundaries have a low electronic activity. Therefore, from all polycrystalline thin-film materials, CIGS cells have the highest efficiency; CIGS has also a high radiation resistance, a property which is promising for space cells. On the other hand, the grain boundary problem of thin-film Si cells might be circumvented by using single crystalline Si in thin-film modules. In addition, CIGS suffers from the rare and scattered resources of the metal In which restricts the production volumes of CIGS to about 1 GW per year. Moreover, even if there were no such restrictions, the production volumes for CIGS modules are still on the 1 MW level. As a consequence, Si wafer technology runs down much faster along the learning curve than CIGS. Thus, it will be difficult for any investor who builds up a large CIGS factory to compete with Si wafer-based technology and to obtain a return of investment.

Only a thin-film technology based on crystalline Si itself might be able to compete with Si wafer technology, which is a moving target. Mass production, together with basic and applied research on low-temperature processing, laser assisted processing etc. will result in further cost reductions for wafered Si.

There are several interesting "third generation" concepts to increase the efficiency of solar cells above 30 %. Most of them are still in the state of basic research, some are even speculative. However, PV still leaves options for improvements.

In any case, PV will be able to contribute a noticeable fraction to German electricity production. However, one can hardly imagine that large-area PV is based on any other material than Si, on the long range assisted by either an appropriate new top material for a tandem cell and/or photonic converters.

## Acknowledgements

The author thanks Ralf Bergmann, Hans-Joachim Queisser, Uwe Rau, Hans-Werner Schock and Markus Schubert for fruitful scientific collaborations. Christopher Berge, Lydia Diegel, Elsette Domaines-Ott, Isabel Kessler, Regine Noack, Fritz Pfisterer and Christine von Rekowski were of indispensable help during the preparation of this manuscript.

## References

1. W. Hoffmann, in *Proc. 17th Europ. Photovolt. Solar Energy Conf.*, edited by B. Mc Nelis, W. Palz, H. A. Ossenbrink, and P. Helm (WIP-Renewable Energies, München, Germany, 2002), p. 851.

2. J. Siemer, *Photon* **1**, 78 (2004).
3. M. Yoshimi, T. Sasaki, T. Suezaki, T. Meguro, K. Matsuda, K. Santo, K. Wadano, A. Nakajima, and K. Yamamoto, High Efficiency Thin Film Silicon Hybrid Solar Cell Module on 1M<sub>2</sub>-Class Large Area Substrate, to be published.
4. W. Shockley and H.-J. Queisser, *J. Appl. Phys.* **32**, 510 (1961).
5. The original prizes in Fig. 2 base on US \$ and consider inflation of the US \$ with respect to the year 2002. The US currency is here converted into EURO by assuming a 1:1 exchange rate. The overall behavior of the curve does not change by this conversion.
6. Maycock, PV Energy Systems, market data based on PV-News, Maycock.
7. In addition there are about 20 GW power available in private or industrial power plants.
8. At present (beginning of 2004), the world-wide production of modules is about 0.7 GW/year, the German production rate is about 0.1 GW/year.
9. L. Frantzis, E. Jones, and A. D. Little, in *Proc. 16 th EU Photovolt. Solar. Energ. Conf.*, edited by H. Scheer, B. McNelis, W. Palz, H. A. Ossenbrink, and P. Helm (James & James, London, UK, 2000), pp. 2100–2103.
10. T. Schott, in *Photovoltaic Guidebook for Decision Makers*, edited by A. Bubenzer and J. Luther (Springer, Berlin, Germany, 2003), pp. 107–170.
11. K. Zweibel, *Solar Energy Materials & Solar Cells* **63**, 375 (2000).
12. H. Schade and H. Maurus, in *Photovoltaic Guidebook for Decision Makers*, edited by A. Bubenzer and J. Luther (Springer, Berlin, Germany, 2003), p. 170.
13. X. Deng and E. A. Schiff, in *Handbook of Photovoltaic Science and Engineering*, edited by A. Luque and S. Hegedus (Wiley, New York, 2003), p. 505.
14. C. Gemmer, Ph.D. thesis, Universität Stuttgart, 2003.
15. X. Wu, J. C. Keane, R. G. Dhere, C. DeHart, D. S. Albin, A. Duda, T. A. Gessert, S. Asher, D. H. Levi, and P. Sheldon in *Proc. 17th Europ. Photovolt. Solar Energy Conf.*, edited by B. McNelis, W. Palz, H. A. Ossenbrink, and P. Helm (WIP-Renewable Energies, München, Germany, 2002), p. 995.
16. B. E. McCandless and J. R. Sites, in *Handbook of Photovoltaic Science and Engineering*, edited by A. Luque and S. Hegedus (Wiley, New York, USA, 2003), p. 617.
17. J. J. Loferski, *J. Appl. Phys.* **27**, 777 (1956).
18. J. Zhao, A. Wang, and M. A. Green, *Prog. Photovolt: Res. Appl.* **7**, 471 (1999).
19. J. H. Werner, J. Mattheis, and U. Rau, in *The Path to Ultra High Efficient Photovoltaics*, edited by A. Jaeger-Waldau (JRC, Ispra, Italy, 2004), (in print).
20. U. Rau and J. H. Werner, *Appl. Phys. Lett.* **84**, 3735 (2004)
21. K. Ramanathan, M. A. Contreras, C. L. Perkins, S. Asher, F. S. Hasoo, J. Keane, D. Young, M. Romero, W. Metzger, R. Noufi, J. Ward, and A. Duda, *Prog. Photovolt. Res. Appl.* **11**, 225 (2003).
22. U. Rau and H.-W. Schock, *Appl. Phys. A* **69**, 131 (1999).
23. W. N. Shafarman and L. Stolt, in *Handbook of Photovoltaic Science and Engineering*, edited by A. Luque and S. Hegedus (Wiley, New York, USA, 2003), p. 567.
24. A. Jasenek and U. Rau, *J. Appl. Phys.* **90**, 650 (2001).
25. C. N. Jardine, G. J. Conibeer, and K. Lane, in *Proc. 17th Europ. Photovolt. Solar Energy Conf.*, edited by B. McNelis, W. Palz, H. A. Ossenbrink, and P. Helm (WIP-Renewable Energies, München, Germany, 2002), p. 724.

26. A. Andersson, *Prog. Photovolt. Res. Appl.* **8**, 61 (2000).
27. R. B. Bergmann, *Appl. Phys. A* **69**, 187 (1999).
28. J. H. Werner, in *Technical Digest of 13th Sunshine workshop on thin film solar cells*, edited by M. Konagai (NEDO, Tokyo, Japan, 2000), pp. 41–48.
29. J. H. Werner, R. Dassow, T. J. Rinke, J. R. Köhler, and R. B. Bergmann, *Thin Solid Films* **383**, 95 (2001).
30. B. Sopori, in *Thin-film Silicon Solar Cells*, edited by A. Luque and S. Hegedus (Wiley, New York, USA, 2003), p. 307.
31. K. Yamamoto, M. Yoshimi, Y. Tawada, Y. Okamoto, and A. Nakajima, *J. Non-Cryst. Solids* **266–269**, 1082 (2000).
32. K. Taretto, U. Rau, and J. H. Werner, *J. Appl. Phys.* **93**, 5447 (2003).
33. R. B. Bergmann, T. J. Rinke, C. Berge, J. Schmidt, and J. H. Werner, *Solar Energy Materials & Solar Cells* **74**, 213 (2002).
34. J. H. Werner, T. A. Wagner, C. Gemmer, C. Berge, W. Brendle, and M. Schubert, in *Proc. 3rd World Conf. on Photovoltaic Solar Energy Conv., Osaka, Japan, 2003* (in print).
35. K. Feldrapp, R. Horbelt, R. Auer, and R. Brendel, *Progr. Photovolt. Res. Appl.* **11**, 105 (2003).
36. P. Verlinden, A. Blakers, K. Weber, J. Babaei, V. Everett, M. Kerr, M. Stuckings, D. Gordeev, and M. Stocks, in *Technical Digest of the International PVSEC-14* (PVSEC Organising Committee, Bangkok, Thailand, 2004), p. 17.
37. A. Blakers, K. Weber, M. Stocks, J. Babaei, V. Everett, M. Kerr, and P. Verlinden, in *13th Workshop on Crystalline Silicon Solar Cell Materials and Processing* (NREL, Denver, 2003), p. 3.
38. A. De Vos, *Endreversible Thermodynamics of Solar Energy Conversion* (Oxford University Press, Oxford, UK, 1992), p. 131.
39. J. H. Werner, S. Kolodinski, and H.-J. Queisser, *Phys. Rev. Lett.* **73**, 3851 (1994).
40. J. H. Werner, R. Brendel, and H.-J. Queisser, in *Proc. 1st World Conf. Photovoltaic Energy Conversion* (IEEE, New York, 1994), p. 1742.
41. M. A. Green, *Third Generation Photovoltaics, Advanced Solar Energy Conversion* (Springer, Berlin, Germany, 2003).
42. A. Marti and A. Luque, *Next Generation Photovoltaics – High Efficiency through Full Spectrum Utilization* (Institute of Physics, Bristol, UK, 2004).
43. R. R. King, C. M. Fetzer, P. C. Colter, K. M. Edmondson, A. A. P. Stavrides, H. Yoon, G. S. Kinsey, H. L. Cotal, J. H. Ermer, R. A. Sherif, and N. H. Karam, in *Proc. 3rd World Conf. Photovoltaic Osaka, Japan, 2003* (in print).
44. R. M. Swanson, in *Handbook of Photovoltaic Science and Engineering*, edited by A. Luque and S. Hegedus (Wiley, New York, USA, 2003), p. 449.
45. G. Güttler and H.-J. Queisser, *Energy Conversion* **10**, 51 (1970).
46. A. Luque and A. Marti, *Phys. Rev. Lett.* **78**, 5014 (1997).
47. A. Marti, L. Cuadra, and A. Luque, in *Next Generation Photovoltaics*, edited by A. Marti and A. Luque (Institute of Physics Publishing, Bristol, UK, 2004), p. 140.
48. P. Würfel, *Solar Energy Mater. Solar Cells* **46**, 43 (1997).
49. T. Trupke, M. A. Green, and P. Würfel, *J. Appl. Phys.* **92**, 1668 (2002).
50. T. Trupke, M. A. Green, and P. Würfel, *J. Appl. Phys.* **92**, 4117 (2002).

## Part II

### Optical Properties

# Optical Transmission through Periodically Nano-structured Metal Films

L. Martín-Moreno<sup>1</sup> and F. J. García-Vidal<sup>2</sup>

<sup>1</sup> Departamento de Física de la Materia Condensada  
ICMA-CSIC, Universidad de Zaragoza, 50009 Zaragoza, Spain

<sup>2</sup> Departamento de Física Teórica de la Materia Condensada  
Universidad Autónoma de Madrid, 28049 Madrid, Spain

**Abstract.** We present a review of some of the recent developments on the use of surface plasmons for subwavelength optics. As example of how surface plasmons are helping us to play new tricks with light, we focus on the optical transmission through *shape periodically* nano-structured metal films.

## 1 Introduction

In the last few years there has been a renewed interest in the optical properties of metallic structures. This interest is driven by the desire to reduce the dimensions of optical devices, which is becoming possible due to technological advances in the micro- and nano- patterning of metals.

The dielectric optical response of a metal is mainly governed by its free electron plasma. Below its plasma frequency, the real part of the dielectric constant ( $\epsilon_M$ ) of a metal is negative, so the wavevector of light propagating inside a metal has an imaginary component and the metal behaves as a photonic insulator. A negative  $\epsilon_M$  has another important consequence: a metal surface may support localized electromagnetic (EM) resonances, called surface plasmons (SPs) [1]. These resonances propagate along the surface of a metal, and can be tailored to concentrate and guide light in very small volumes. For optical frequencies, SPs appearing in good metals (like silver) may propagate distances of the order of  $10 - 100 \mu\text{m}$  before being absorbed, going up to  $1 \text{ mm}$  in the near infrared. Therefore, if we are interested in highly integrated planar optical devices smaller than these typical propagation distances, the inherent absorption present in conductive systems could be not a too serious drawback, and SPs could be considered as a promising route to sub-wavelength optics [2].

In the first part of this paper we overview some of the recent developments on the use of SPs for subwavelength optics. As example of these new capabilities, in the second part of the paper we focus on the optical transmission through *shape periodically* nano-structured metal films.

## 2 SPs as a Route to Sub-wavelength Optics

The dispersion relation of a SP propagating in a metal-vacuum interface is given by the following equation

$$k(\omega) = \frac{w}{c} \sqrt{\frac{\epsilon_M(\omega)}{1 + \epsilon_M(\omega)}} \quad (1)$$

where  $\omega$  is the frequency,  $c$  is the speed of light and  $\epsilon_M(\omega)$ , the frequency-dependent dielectric constant of the metal.

As its dispersion curve lies outside the light-cone, a SP on a metal surface can not be excited by an incident plane wave. This is the main problem when dealing with these type of surface EM resonances: the coupling of light into SPs and the reversed process, coupling SPs to light. There are three main different strategies to overcome this difficulty.

The most common technique to excite SPs consists in evaporating a thin metal film onto an optically dense substrate. Illuminating through the substrate at an angle of incidence slightly larger than the critical angle allows the wave vector of the incident light to match that of the SP. This technique, known as attenuated total reflection (ATR) [3], has been used recently to analyze the propagation of SPs on thin metal stripes [4]. These metal stripes could be used in optoelectronic devices as a common channel to propagation of both electrical current and light via SPs.

The second approach consists in launching SPs by focusing a laser beam onto a topological defect [5]. Combining this technique with fluorescence imaging, it has been possible to test the first realizations of a SP-based Bragg mirror and a SP-based beam-splitter [6]. These promising results could open the possibility of using SPs in order to create devices able to do sub-wavelength optics in two dimensions.

The third way for coupling light into SPs is by creating a periodic modulation of the metal surface. This modulation provokes the folding of the SP bands inside the first Brillouin zone, i.e., inside the light cone. Already in 1902, Wood [7] reported anomalous behaviour in the diffraction of light by metallic gratings that we now know were due to the excitation of SPs. Periodicity can also lead to the opening of a full photonic band gap (PBG) in the SP spectrum, in a way very much similar to the opening of a photonic band gap in photonic crystals. By analyzing the reflectance spectrum of an hexagonal array of photoresist dots on a glass substrate coated with a thin film of silver, Kitson *et al.* [8] were able to report the appearance of a SP-PBG between 1.91 and 2.0 eV, within the visible range of EM spectrum. The existence of a band gap in which light propagation is forbidden can lead to many optical applications. Among others, it is possible to guide light in two-dimensions by just creating some line defects into the periodic array: light with the appropriate energy inside the gap can not couple to bulk crystal EM modes and its propagation is confined inside the line defect. Experimental realization of

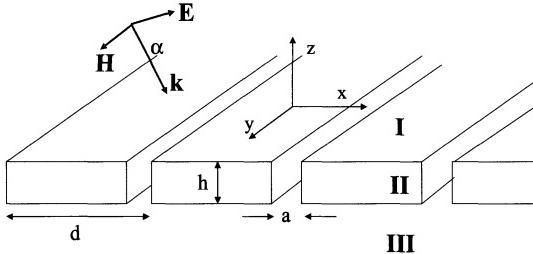
these ideas has been recently reported [9], rendering the prospect of photonic circuits based on SPs.

### 3 Optical Transmission Through Nano-structured Metals

In 1998, Ebbesen et al. [10] found experimentally that the transmission of light through subwavelength hole arrays made in a metal film may be, for some particular resonant wavelengths, orders of magnitude larger than expected if the holes were acting *shape independently*. Standard aperture theory stated that, for a single hole, the transmittance ( $T$ ), is roughly  $T \sim a^2 (a/\lambda)^4$  ( $a$  being the hole radius) [11], *shape i.e.* due to the wave character of the light the normalized-to-area transmittance through a hole is very small whenever  $a \ll \lambda$ . Furthermore, that theory was done for the case of a metal film with infinitesimal thickness; for the realistic case of finite metal thickness the transmittance should be even smaller as, in the subwavelength limit, all EM modes inside the waveguide are evanescent. It is worth pointing out that, before Ebbesen's experiment, hole arrays had been extensively studied, due mainly to their properties as selective filters. For this, the high-pass filtering properties are provided by the hole EM cutoff, while the low-pass filter is due to the redistribution of energy caused by the periodic array when a new diffraction order becomes propagating. More precisely, denoting the hole cut-off wavelength by  $\lambda_c$ , and the lattice parameter by  $d$ , hole arrays were known to act as band-pass filters for  $d < \lambda < \lambda_c$ . This property has been studied in several frequency regimes. as microwave [12], far infrared [13], mid infrared [14] and infrared [15]. However, the extraordinary optical transmission (EOT) discovered by Ebbesen et al. presented two main differences with previous works. First, experiments were performed in the optical regime, and second, the geometrical parameters defining the structure were such that  $\lambda_c < d < \lambda$ ; this parameter range had not, up to our knowledge, been studied before, perhaps because nothing remarkable was expected for wavelengths beyond cut-off. Although already the first experiments [10] showed that EOT was related to the excitation of SPs, the transmission mechanism was unclear. The initial theoretical efforts concentrated on the simpler system of a periodic array of 1D apertures (slits), with realistic calculations for the 2D Ebbesen's geometry appearing shortly afterwards. Here we review the basics of the enhanced transmission phenomena for both the 1D- and 2D-array of apertures in an opaque metal film.

#### 3.1 1D Slit Array

Consider the transmission grating (TG) depicted in Fig. 1 (find there also the definition for the different geometrical parameters defining the array and the choice of axis).



**Fig. 1.** Scheme of a 1D array of slits in a metal film. Regions I and III are filled by a dielectric materials, that here we take to be air in both cases

This system is reminiscent of reflection gratings, in which case region III in Fig. 1 is filled with metal. Reflection gratings have been extensively studied in the past. As commented in the introduction, Wood [7] found remarkable absorption anomalies in the reflectance spectra of smooth metallic gratings illuminated by p-polarized light (E-field perpendicular to the grating axis). Fano [16] associated these anomalies to the excitation of SPs. Later on, Hessel and Oliner found [17] another type of absorption anomaly, appearing in deep gratings, associated to the formation of standing waves inside the grooves. The small volume of these modes results in strong EM field enhancement, so these systems are excellent substrates for surface-enhanced Raman studies [20].

Apart from some theoretical works [18,19], TGs had not received so much attention, but after Ebbesen's experiment they have been thoroughly studied. TGs have proved interesting in their own right, and some of the basics of the enhanced transmission phenomena are already present in these systems. However it must be kept in mind that in a slit waveguide there is always a propagating mode inside the waveguide, no matter how narrow it is, while a hole waveguide present a cutoff frequency and all modes are evanescent for hole diameters smaller than (roughly) half a wavelength. Therefore, wave propagation is radically different for slits and holes.

### 3.1.1 Theoretical Formalism

Let us concentrate on the optical transmission properties of arrays of sub-wavelength slits, for p-polarized light. For s-polarization (E-field along the slit axis) the transmission does not present such a rich resonant behavior, and will not be discussed here. There are different theoretical frameworks that give virtually exact results for the transmittance of such a simple structure. However, the physical mechanisms responsible for the transmittance spectra are best captured by a simplified quasi-analytical model, where fields are represented in different regions by a modal expansion [21]. We have previously demonstrated [22] the accuracy of this theoretical framework in comparison with, for example, transfer matrix methods. Two main approximations are considered within our modal expansion:

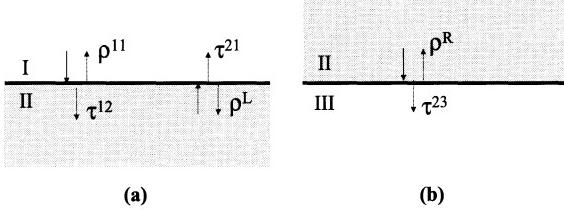
- only the fundamental eigenmode in the modal expansion of the EM fields inside each slit is considered, which is justified because in the subwavelength regime the fundamental mode is the only propagating one, and dominates the transmittance
- surface impedance boundary conditions (SIBC) [23] are imposed on the metal dielectric interfaces. The approximation of SIBC is applicable when the skin depth in the metal (of the order of  $20\text{ nm}$ , for good metals in the optical regime) is much smaller than all other length scales in the system. Strictly, this is not the case, as we are going to consider slit widths of a few tens of nanometers. However, as the propagation constant of the fundamental mode of the slit does not depend on slit width, it turns out that the SIBC is a good approximation for this kind of systems even in the optical regime. Moreover, given that in the considered structures the "horizontal" metal-dielectric interfaces are much larger than the "vertical" ones, we further simplify the model (just for computational convenience) by assuming that the vertical walls of the slits are perfect metal surfaces.

For p-polarization, the modal expansion for the magnetic field is

$$\begin{aligned} H_y(x, z) &= \Psi_0^+(x, z) + \sum_n r_n \Psi_n^-(x, z), \quad \text{for } z \text{ in region I} \\ H_y(x, z) &= \sum_m \{ A_m \Phi_m^+(x, z) + B_m \Phi_m^-(x, z) \}, \quad \text{for } z \text{ in region II} \\ H_y(x, z) &= \sum_n t_n \Psi_n^+(x, z), \quad \text{for } z \text{ in region III}. \end{aligned} \tag{2}$$

where, for an incident wave impinging at angle  $\alpha$  with the normal,  $k = \omega/c$ ,  $k_{xn} = k \cos(\alpha) + (2\pi/d)n$ ,  $k_{zn} = \sqrt{k^2 - k_{xn}^2}$ ,  $\Phi_m^\pm(x, z) = \exp(\pm ikz)/\sqrt{d}$  if  $x \in [md - a/2, md + a/2]$  and zero otherwise, and  $\Psi_n^\pm(x, z) = \exp(ik_{xn}x \pm ik_{zn}z)/\sqrt{d}$ . In Eq. 2, all summations go from  $-\infty$  to  $+\infty$ . The electric field components  $E_x$  and  $E_z$ , can be readily calculated via the Maxwell equation  $\nabla \times \mathbf{E} = ik\mathbf{H}$ .

Calculating the transmission and reflection coefficients  $t_n$  and  $r_n$  into the different diffraction orders, and the slit amplitudes  $A_m$  and  $B_m$  is a simple exercise of matching fields at both I-II and II-III interfaces. However, it is convenient to extract the scattering coefficients in the three-region system from the scattering coefficients of two independent two-region systems: the I-II and the II-III systems, in which all regions are taken as semi-infinite (even region II). Let us concentrate on the zero-order transmission amplitude,  $t_0$ ; expressions for all other scattering amplitudes can be calculated trivially in a similar manner. We define (see Fig. 2) the scattering coefficients in the two-region systems as follows: in the I-II system, an incident plane wave coming from region I reflects, when reaching I-II interface, with probability amplitude  $\rho^{11}$  or transmits through the interface with amplitude  $\tau^{12}$ . When approaching the II-I interface coming from region II, the slit fundamental eigenmode



**Fig. 2.** Definitions of the different scattering coefficients of the two different two-region systems appearing in transmission gratings

either reflects with amplitude  $\rho^L$ , or transmits to region I with amplitude  $\tau^{21}$ . Actually,  $\rho^{11}$  and  $\tau^{21}$  would be matrices, given that the final state could be any of the diffraction orders. In the II-III system, the propagating mode coming from region II bounces back at the II-III interface with amplitude  $\rho^R$  or transmits to region III with amplitude  $\tau^{23}$ .

Knowing all these coefficients, the scattering coefficients for the real three-region system can be easily calculated summing up all multiple scattering processes, the only additional ingredient needed is the phase accumulated by the EM field when propagating inside the waveguide,  $\phi_P = \exp(i\theta_P)$ , with  $\theta_P = kh$ . The final result is

$$t_0 = \frac{\tau^{12} \phi_P \tau^{23}}{1 - \rho^L \rho^R \phi_P^2} \quad (3)$$

The expressions for the different two-region coefficients can be found in [22]. Let us just recall that

$$\rho^L = \rho^R = -\frac{1 - (1 + Z_s) f}{1 + (1 - Z_s) f} \quad (4)$$

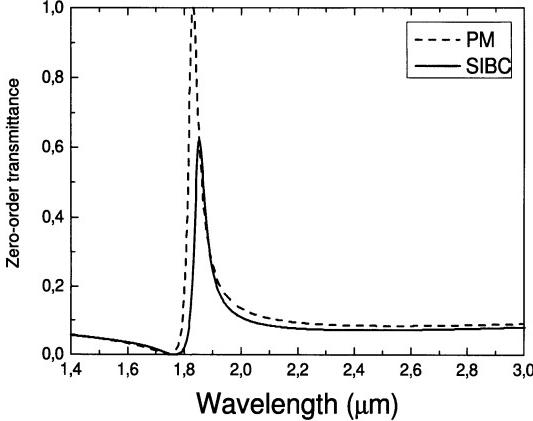
where  $Z_s = [\epsilon(\omega)]^{-1/2}$  is the metal surface impedance and

$$f = \sum_{n=-\infty}^{n=+\infty} \frac{[\text{sinc}(k_{xn}a/2)]^2}{k_{zn}/k + Z_s} \quad (5)$$

where  $\text{sinc}(x) = \sin(x)/x$ , and  $f$  is a quantity that plays a central role in all scattering coefficients. A crucial point, which strongly influences the transmission (and reflection) properties is that  $f$  is singular whenever any of the diffraction orders satisfy  $k_{zn} = -kZ_s$ , which is the condition for existence of a SP on a *shape flat* metal surface (within the SIBC approximation).

### 3.1.2 Results

Let us consider a slit array made in silver film [24] with  $d = 1.75 \mu\text{m}$ ,  $a = 0.30 \mu\text{m}$  and  $W = 0.4 \mu\text{m}$ , which are typical experimental parameters [25], illuminated by an normal-incident p-polarized EM field. We concentrate in



**Fig. 3.** Comparison between  $T(\lambda)$  calculated with our simplified modal expansion for surface impedance (full curve) and perfect metal (dashed curve) boundary conditions. The geometrical parameters are:  $d = 1.75 \mu\text{m}$ ,  $a = 0.3 \mu\text{m}$  and  $W = 0.4 \mu\text{m}$  for normal-incident p-polarized light

the spectral region  $\lambda > d$ , where only the zero-order transmission coefficient  $t_0$  contributes to the total current.

Figure 3 renders the comparison between the transmittance spectra calculated with the simplified modal expansion previously described, and the calculation in which SIBC are substituted by perfect metal (PM) boundary conditions.

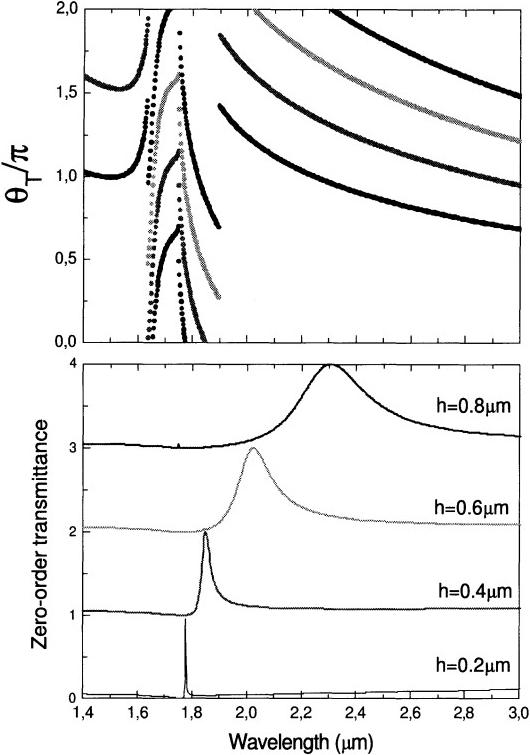
As Fig. 3 shows, the transmittance spectrum presents a peak, where light is almost 100% transmitted for the PM case. This is reduced by around 40% when the absorption in the metal is included.

In order to gain physical insight into the origin of this transmission peak, it is convenient to rewrite  $\rho^L = R \exp(i\theta_S)$ . Then, defining  $\theta_T = 2(\theta_P + \theta_S)$ ,  $t_0$  can be written as

$$t_0 = \frac{\tau^{12} \phi_P \tau^{23}}{1 - R^2 \exp(i\theta_T)} \quad (6)$$

For narrow slits, due to the large impedance mismatch,  $R \approx 1$  (as the slit waveguide mode is propagating, current conservation forces  $R$  to range from 0 to 1), so the transmittance is dominated by the total phase  $\theta_T$ . Therefore, the system behaves as a Fabry-Perot interferometer: whenever  $\theta_T$  in an integer times  $2\pi$ , all partial waves in the multiple scattering series interfere constructively in the forward direction, leading to a maximum in  $T(\lambda)$ . However, as a difference with a standard Fabry-Perot interferometer, added to the wavelength dependence of the optical-path phase  $\theta_P$ , the scattering phase ( $\theta_S$ ) varies strongly close to the SPs spectral locations. This is illustrated in Fig. 4, showing  $\theta_T$  and  $T$  as a function of wavelength, for different metal thicknesses.

Figure 4 shows a series of transmittance peaks, reaching 100% transmission if absorption is neglected. Notice that the slit only occupies a small fraction of the device unit cell, so the structured metal surface is acting as a funnel, collecting all light impinging on the surface into a very narrow



**Fig. 4.** Total phase  $\theta_T$  (panel (a)) and zero-order transmittance (panel (b)) as a function of  $\lambda$ , for a normal-incident p-polarized plane-wave impinging on a slit array with  $d = 1.75 \mu\text{m}$ ,  $a = 0.3 \mu\text{m}$  for different thicknesses:  $W = 0.2, 0.4, 0.6$  and  $0.8 \mu\text{m}$ . Each transmittance curve has been shifted by +1 with respect to the previous one for better visualization

aperture. These resonances present large differences in peak linewidths (and, correspondingly, the sensitivity to absorption). It can be shown that the resonance linewidth is inversely proportional to the derivative of  $\theta_T$  with respect to  $\lambda$ , which is large (due to the behavior of  $\theta_S$  whenever there is a surface resonance in the system). Also the EM field distribution changes from being mostly the one corresponding to two coupled SPs, to having a cavity mode character (reminiscent of the open-organ pipe modes), depending on whether the peak spectral position is close to the  $k_{zn} = kZ_s$  condition or away from it, respectively [22].

### 3.2 2D Hole Array

As previously stated, the studies of slit arrays are of limited value for understanding the transmission through subwavelength hole arrays, given that a subwavelength hole does not support a propagating TEM mode, while a subwavelength slit does. Computation of the transmittance for hole arrays is much more computationally demanding than for slit arrays, due to the lower symmetry in the former system. Popov et al. [26] presented a calculation based on a modal expansion taking into account the spectral dependence of the dielectric constant in the metal,  $\epsilon_M(\omega)$ . Their calculations showed transmission peaks, which were associated to an EM mode propagating along the

holes, that appeared in the hole array when a realistic  $\epsilon_M(\omega)$  was taken into account, and was not present when the metal was treated as perfectly conducting. On the contrary, a calculation performed with a modal expansion plus the approximations previously described for the slit array calculation, reported EOT, *shape even in the metal were perfectly conducting* [27]. Here we adhere to this last view, and present what we believe are the basics of the EOT phenomena in 2D hole arrays, by studying for a highly simplified model. This model treats the metal as a perfect conductor, showing that EOT is possible even when there is no propagating mode inside the holes.

### 3.2.1 Minimal Model

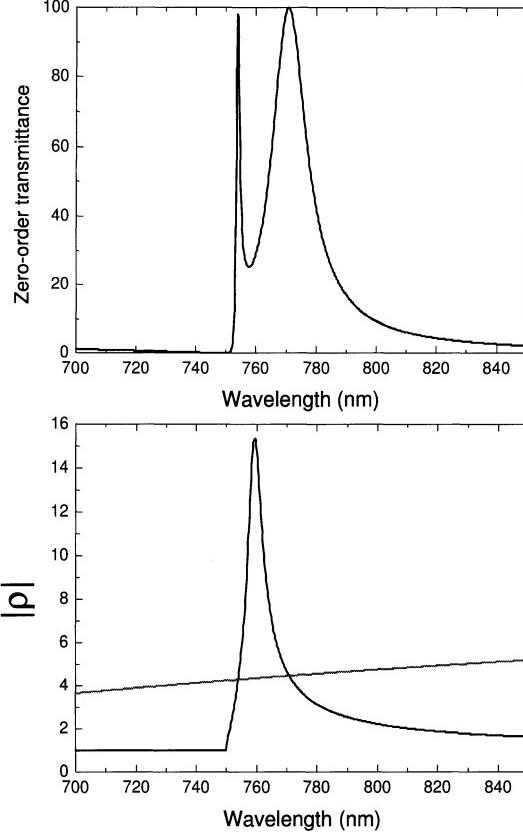
The calculation for the transmittance through an square array of square holes, within the approximations described in the subsection 3.1.1, was presented in [27]. In this case, all diffraction orders (resulting from a change in the incident parallel momentum by a reciprocal lattice vector), as well as both polarizations must, in principle, be taken into account. However, a minimal model, which considers just the incoming wave plus two diffraction orders (for a normal-incident linearly-polarized incoming wave, with the E-field pointing along the  $x$ -direction, we consider the p-polarized diffraction order with  $k_x = 0, \pm 1$  times  $2\pi/d$ ) provides an excellent approximation to the transmittance spectra. In this paper, we also treat the metal as perfectly conducting ( $Z_s = 0$ ). This approximation further simplifies the calculation; considering absorption in the metal would reduce the transmission peaks [27] but would not alter the physical picture [29]. Figure 5 shows the comparison between the  $T(\lambda)$ , for a square array of square holes [30] calculated with this minimal model.

Clearly, the minimal model captures the physics of the EOT, with the advantage that it can be analytically worked out. In this case, the multiple scattering series (3) gives

$$t_{00} = \frac{\tau^{12} e^{-|q|W} \tau^{23}}{1 - \rho^2 e^{-2|q|W}} \quad (7)$$

where, denoting  $q = \sqrt{k^2 - (\pi/a)^2}$ ,  $k_{z1} = \sqrt{k^2 - k_{x1}^2}$ ,  $S_0 = 2\sqrt{2}a/(\pi d)$ ,  $S_1 = S_0 \text{sinc}[k_{x1}d/2]$ ,  $G_1 = S_0^2 + 2S_1^2 k/k_{z1}$  and  $G_2 = q/k$ , it is found that  $\tau^{12} = 2S_0/(G_2 + G_1)$ ,  $\tau^{23} = 2G_2/(G_2 + G_1)$ , and  $\rho = (G_2 - G_1)/(G_2 + G_1)$ .

The appearance of EOT is related to resonant denominators in Eq. (7), which are possible because, for evanescent modes, current conservation only forces  $\text{Im}(\rho) > 0$ , but says noting about the modulus of  $\rho$ . This point is illustrated in Fig. 5, showing both real and imaginary parts of  $\rho(\lambda)$ . As  $\rho$  is a causal function, its real and imaginary part must satisfy Kramers-Kronig relations; furthermore, the peak in  $\text{Im}(\rho)$  signals the presence of a localized surface EM mode in the system (the peak in  $|\rho|$  in Fig. 5 is due to a peak in  $\text{Im}(\rho)$ ). The finite spectral width of this peak is inversely proportional to



**Fig. 5.** Upper panel:  $T(\lambda)$ , for a square array of holes in a perfect metal, calculated within the minimal model, for normal-incident p-polarized light. Lower panel:  $|\rho(\lambda)|$  (black curve) and  $e^{|q|W}$ . When this two magnitudes are equal, there is a peak in  $T(\lambda)$ . We consider  $d=750$  nm,  $W=200$  nm and square holes with side length  $a=284$  nm (in order to mimic circular holes with a diameter of 320 nm)

the time that the EM takes to leave the system due to radiation losses ( $t_{\text{rad}}$ ), so more precisely speaking, what the system develops is a leaky surface EM mode. There are two such modes (one for each metal-dielectric interface), which are coupled through the evanescent fields inside the holes. This results in symmetric and antisymmetric combinations, in much the same way electronic states of isolated atoms combine to form molecular levels [27]. The case of transmission through holes in a metal film resting on a dielectric substrate was described in [28] and, in this picture, it would correspond to the formation of surface plasmon "molecular levels" (SPML) from non-equal surface plasmon "atoms". This "surface plasmon molecular states" are coupled to radiation modes, and EOT occurs via them. In this sense it can be stated that EOT is the physics of two localized levels coupled to a continuum. Additional insight can be gained if we calculate the SPML spectral position neglecting the coupling to the radiative modes. In this case, from the frequency difference between the position of this modes is related to the time the photon has to stay in the system in order to "see" the resonant levels,  $t_{\text{res}}$ . Notice that  $t_{\text{res}}$  depends on the coupling between SPs, i.e., on metal thickness, while  $t_{\text{rad}}$  does not, as it is a characteristic of a single surface. Different regimes

of tunnelling (sequential for  $t_{\text{rad}} < t_{\text{res}}$  and resonant) in the opposite case were predicted [27] (see also Ref. [31] for a description of similar physics in resonant tunnelling in quantum mechanics) and measured subsequently [32].

### 3.3 Single Apertures Flanked by Corrugations

We would like to end with a short note on non-periodic systems. As was discussed, surface EM waves were at the heart of the EOT phenomena. However, it was reasonable to think that perfect periodicity was not strictly necessary for the existence of surface waves. This opens the possibility of using SPs in order to obtain EOT through a *shape single aperture*, effect that has now been studied both experimentally [33] and theoretically [34], if the metal surface is corrugated in the side the light is impinging on. Moreover, it has been found that very strong directional emission is possible through single apertures if the corrugation is on the exit side [35,36,37]. Recently, light demultiplexing [38] and focusing effects [39] have been reported in single apertures in corrugated metals. It might well be that, although based on guiding ideas found when studying periodic systems, many surprises and applications are still to be found in finite non-periodic systems.

### Acknowledgements

We would like to acknowledge financial support by the Spanish MCyT through contracts MAT2002-01534 and MAT2002-00139, and fruitful discussions and collaborations with J. Bravo-Abad, T.W. Ebbesen, H.J. Lezec, J.A. Porto, T. Thio and J.B. Pendry.

### References

1. R.H. Ritchie, Phys. Rev. **106**, 874 (1957).
2. W.L. Barnes, A. Dereux, and T.W. Ebbesen, Nature **424**, 824 (2003).
3. E. Kretschmann and H. Raether, Z. Naturforsch. A **23**, 2135-2136 (1968); A. Otto, Z. Phys. **216**, 398 (1968).
4. J.C. Weeber *et al.*, Phys. Rev. B **64**, 0455411 (2001).
5. B. Hecht, H. Bielefeldt, L. Novotny, Y. Inouye, and D.W. Pohl, Phys. Rev. Lett. **77**, 1889 (1996).
6. H. Ditlbacher, J.R. Krenn, G. Schider, A. Leitner, and F.R. Aussenegg, Appl. Phys. Lett. **81**, 1762 (2002).
7. R.W. Wood, Phil. Mag. **4**, 396 (1902).
8. S.C. Kitson, W.L. Barnes, and J.R. Sambles, Phys. Rev. Lett. **77**, 2670 (1996).
9. S.I. Bozhevolnyi, J. Erland, K. Leosson, P.M.W. Skovgaard, and J.M. Hvam, Phys. Rev. Lett. **86**, 3008 (2001); S.I. Bozhevolnyi, V.S. Volkov, K. Leosson, and A. Boltasseva, Appl. Phys. Lett. **79**, 1076 (2001).
10. T.W. Ebbesen, H.J. Lezec, H.F. Ghaemi, T. Thio, and P.A. Wolff, Nature **391**, 667 (1998).

11. H. A. Bethe, Phys. Rev. **66**, 163 (1944).
12. F. Keilmann, Int. Journ. Infrared and Millimeter Waves **2**, 259 (1981).
13. A. Mitsuishi, Y. Otsuka, S. Fujita and H. Yoshinaga, Jap. J. Appl. Phys., **2**, 574 (1963).
14. R. Ulrich, Infrared Phys., **7**, 27 (1967).
15. C.M. Rhoads, E.K. Damon, and B.A. Munk, Appl. Opt. **21**, 2814 (1982).
16. U. Fano, J. Opt. Soc. Am. **31**, 213 (1941).
17. A. Hessel and A.A. Oliner, Appl. Opt. **4**, 1275 (1965).
18. H. Lochbihler, Phys. Rev. B **50**, 4795 (1994).
19. J.J. Kuta *et al.*, J. Opt. Soc. Am. A **12**, 1118 (1995).
20. F.J. Garcia-Vidal and J.B. Pendry, Phys. Rev. Lett. **77**, 1163 (1996).
21. P. Sheng, R.S. Stepleman, and P.N. Sanda, Phys. Rev. B **26**, 2907 (1982).
22. F.J. Garcia-Vidal and L. Martin-Moreno, Phys. Rev. B **66**, 155412 (2002).
23. L.D.Landau, E.M. Lifshitz and L.P. Pitaevskii, *Electrodynamics of Continuous Media*, Pergamon Press, second edition, (1994).
24. We take the dielectric constant for silver from: *Handbook of Optical Constants of Solids*, edited by E.D. Palik (Academic, Orlando, 1985).
25. A. Barbara, P. Quemerais, E. Bustarret, and T. Lopez-Rios, Phys. Rev. B **66**, 161403 (2002).
26. E. Popov, M. Nevière, S. Enoch, and R. Reinisch, Phys. Rev. B **62**, 16100 (2000).
27. L. Martin-Moreno, F.J. Garcia-Vidal, H.J. Lezec, K.M. Pellerin, T.Thio, J.B. Pendry, and T.W. Ebbesen, Phys. Rev. Lett. **86**, 1114 (2001).
28. A. Krishnan *et al.*, Opt. Comm. **200**, 1 (2001).
29. The effect of the field penetration inside the metal can be approximately taken into account by adding a skin depth to the hole lateral dimensions.
30. The choice of square holes is motivated by analitycal convenience. We have checked that an array of circular subwavelength holes present very similar transmittance spectra.
31. A. A. Abrikosov, Phys. Rev. B **66**, 212501 (2002).
32. A. Degiron, H. J. Lezec, W. L. Barnes, and T. W. Ebbesen, Appl. Phys. Lett. **81**, 4327 (2002).
33. T. Thio et al. Opt. Lett. **26**, 1972 (2001).
34. F. J. García-Vidal, H. J. Lezec, T. W. Ebbesen, and L. Martín-Moreno, Phys. Rev. Lett. **90**, 213901 (2003).
35. H. J. Lezec, A. Degiron, E. Deveaux, R.A. Linke, L.Martín-Moreno, F.J. García-Vidal and T.E. Ebbesen, Science **297**, 822 (2002).
36. A.P. Hibbins, J.R. Sambles and C.R. Lawrence, Appl. Phys. Lett. **81**, 4661 (2002).
37. L. Martín-Moreno, F. J. García-Vidal, H. J. Lezec, A. Degiron, and T. W. Ebbesen, Phys. Rev. Lett. **90**, 167401 (2003).
38. J. Bravo-Abad, F. J. García-Vidal and L. Martín-Moreno, Photonics and Nanostruct. **1**, 55 (2003).
39. F. J. García-Vidal, L. Martín-Moreno, H. J. Lezec and T. W. Ebbesen, Appl. Phys. Lett. **83**, 4500 (2003).

# Charged Excitons in the Quantum Hall Regime: Optical Probe of Fractionally Charged Quasiholes

Christian Schüller<sup>1</sup>, Kay-Birger Broocks<sup>1</sup>, Patrick Schröter<sup>1</sup>,  
Christian Heyn<sup>1</sup>, Detlef Heitmann<sup>1</sup>, Max Bichler<sup>2</sup>, Werner Wegscheider<sup>3</sup>,  
Vladimir M. Apalkov<sup>4</sup>, and Tapash Chakraborty<sup>5</sup>

<sup>1</sup> Institut für Angewandte Physik und Zentrum für Mikrostrukturforschung  
Universität Hamburg, Jungiusstraße 11, 20355 Hamburg, Germany  
[schueller@physnet.uni-hamburg.de](mailto:schueller@physnet.uni-hamburg.de)

<sup>2</sup> Walter-Schottky-Institut der Technischen Universität München  
Am Coulombwall, 85748 Garching, Germany

<sup>3</sup> Institut für Experimentelle und Angewandte Physik, Universität Regensburg  
93040 Regensburg, Germany

<sup>4</sup> Physics Department, University of Utah  
Salt Lake City, Utah 84112-0830, USA

<sup>5</sup> Department of Physics and Astronomy, University of Manitoba  
Winnipeg, Manitoba Canada R3T 2N2  
[tapash@physics.umanitoba.ca](mailto:tapash@physics.umanitoba.ca)

**Abstract.** We review recent optical experiments on dilute two-dimensional electron systems (2DES) at very low temperatures ( $T < 0.1$  K) and high magnetic fields. In photoluminescence experiments on a 2DES subjected to a quantizing magnetic field around  $\nu = 1/3$ , we have observed an anomalous dispersion of the charged excitons. We have found that the anomaly exists only at a very low temperature (0.1 K) and an intermediate electron density ( $0.9 \times 10^{11} \text{ cm}^{-2}$ ). It is explained to occur due to the perturbation of the incompressible liquid at  $\nu = 1/3$ . The perturbation is induced by the close proximity of a localized charged exciton which creates a fractionally-charged quasi-hole in the liquid. The intriguing experimentally observed puzzle that the anomaly (2 meV) can be destroyed by applying a small thermal energy of  $\sim 0.2$  meV is thereby resolved, as this energy is enough to close the quasi-hole energy gap. This work presents a first ever probe of the quasi-hole gap in a quantum Hall system.

## 1 Introduction

Since its discovery [1], the fractional quantum Hall effect (FQHE), which is a direct manifestation of many-particle interactions in a two-dimensional electron system (2DES), has attracted great and sustaining research interest. The past decades have shown that, in addition to transport experiments, optical experiments like photoluminescence (PL) spectroscopy can provide further and in some sense complementary information about the ground state of quantum Hall systems. In fact, in optical experiments one is able to probe

the bulk of a quantum Hall system, whereas transport is mostly governed by edge channels. However, one of the striking differences of optical experiments, compared to transport, is the presence of photoexcited holes in the system, which can significantly influence its properties.

The optical investigations reported so far can roughly be divided into four categories by the systems under investigation: (i) single heterojunctions (e.g., Ref. [2]) (ii) remotely acceptor-doped heterojunctions (e.g., Ref. [3,4]), and (iii) single quantum wells (e.g., Ref. [5,6,7,8,9,10]). In particular in experiments on quantum wells, which are also the topic of this work, electrons and photoexcited holes reside essentially in the same spatial layer – the quantum well – and Coulomb interaction will be of major importance. Though it was previously shown that in symmetric systems in the high magnetic field limit the electron-electron and electron-hole interactions cancel exactly due to a hidden symmetry [11,12,13], for most of the real systems this symmetry is violated. The role of Coulomb interaction between electrons and holes in optical experiments has been explored in recent years, in particular, in experiments on very dilute 2DES, in the density range  $\sim 10^{10} \text{ cm}^{-2}$ . Here, it was found that the ground state is formed by negatively charged excitons, a bound state of two electrons and one hole. In a magnetic field, the two electrons can form either a singlet or a triplet state; the excitons are then called singlet ( $X_s^-$ ) or triplet ( $X_t^-$ ) excitons, respectively (see, e.g., [14,15,16,17,18,19,20,21,22]). For samples with relatively strong disorder, the commonly accepted picture is that at low density the 2DES breaks up into areas with finite density (electron puddles) and completely depleted regions, where neutral excitons ( $X^0$ ) can form. On the other hand, experiments on high-quality GaAs-AlGaAs samples suggested [23] that charged and neutral excitons may also reside in the same spatial regions. This is, however, controversially discussed, since, theoretically, neutral excitons should not be stable if charged excitons exist as bound states at lower energies. Furthermore, time-resolved experiments on p-doped CdTe-CdMgZnTe quantum wells [24] revealed that, both the positively charged excitons and the residual holes are localized, whereas the neutral excitons remain freely moving. We will show below that the effect of disorder is one of the main ingredients in our experiments, too.

At higher densities, in the range  $\sim 10^{11} \text{ cm}^{-2}$ , the Coulomb interaction between electrons and holes is mostly screened. Naturally, the crossover between these two regimes should reveal interesting interaction related effects. Indeed, in our work we could show that the charged excitons can, under specific conditions, be used as a tool to investigate the properties of a highly correlated quantum Hall liquid [10]. We will show that the interaction between a localized charged exciton and the incompressible liquid in the  $\nu = 1/3$  state leads to the creation of a fractionally-charged hole in the liquid. This low-lying charged excitation is a fundamental consequence of the incompressible nature of the ground state – the liquid state proposed by Laughlin [25].

In addition to the ground state energy, the quasi-hole creation energy for the state proposed by Laughlin is also known very accurately [26]. The energy gap for a well-separated quasiparticle – quasi-hole pair has been derived from low-temperature measurements [27], and theoretically investigated by many authors [28,29]. A few years ago, several groups reported their successful attempts to detect the existence of fractionally charged objects in the edge states of a two-dimensional electron system invoking a chiral Luttinger liquid theory [30]. In our work, we have demonstrated that the lowest-energy charged excitations, the quasiholes, can be probed in an incompressible liquid at  $\nu = 1/3$  (the Laughlin state) via optical spectroscopy [10].

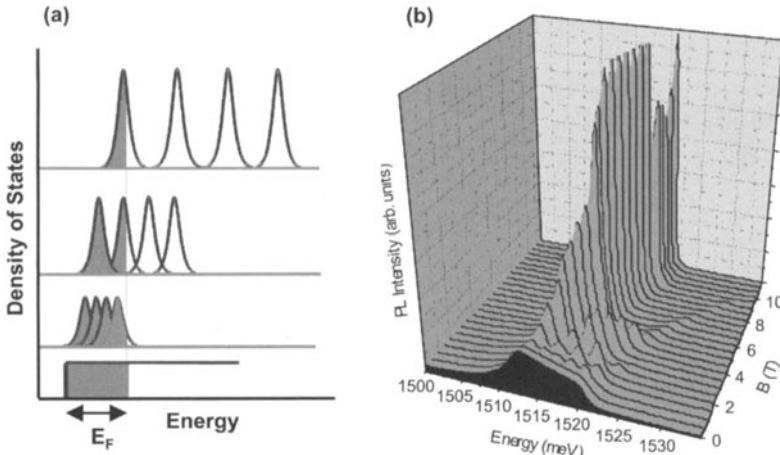
## 2 Magneto-photoluminescence and Absorption in the Quantum Hall Regime

### 2.1 Experimental Details

The samples which were investigated in this work are one-sided modulation-doped GaAs-Al<sub>0.33</sub>Ga<sub>0.67</sub>As single quantum wells with a 25 nm wide GaAs quantum well where a 2DES is formed. One side of the quantum well, either a 70 nm or a 88 nm AlGaAs spacer separates the doped barrier region from the quantum well. On the other side, a 100 period AlGaAs/GaAs (10 nm/3 nm per period) superlattice was grown. On top of the sample, a 7 nm thick Titanium gate was deposited. The sample was glued upside down on a glass substrate using an UV curing optical adhesive. Subsequently, the sample was thinned from the back side by a selective etching process [31] down to the superlattice to a total thickness of about 1.3  $\mu\text{m}$ . By applying a negative gate voltage, we can tune the density in the range between about  $1 \times 10^{10} \text{ cm}^{-2}$  and  $2 \times 10^{11} \text{ cm}^{-2}$ . PL and absorption measurements were performed either in an optical split-coil cryostat, or via glass fibers in a <sup>3</sup>He/<sup>4</sup>He dilution cryostat at a base temperature of  $T = 40 \text{ mK}$  and magnetic fields up to 16 T. In the dilution cryostat a sensor at the sample position indicated that during illumination the temperature directly at the sample is about  $T = 100 \text{ mK}$ , while the base temperature is still  $T = 40 \text{ mK}$ . Circularly polarized light was created directly inside the mixing chamber and left and right circularly polarized spectra were measured by ramping the magnet from positive to negative magnetic fields. For the PL, the sample was excited by a Ti:Sapphire laser at 750 nm. For the absorption measurement, a white light source was used.

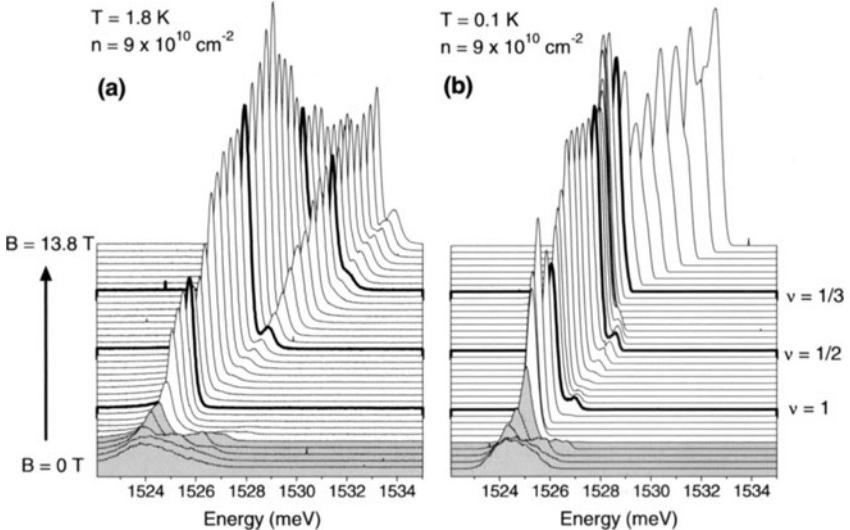
### 2.2 High Electron Density

If a perpendicular magnetic field is applied to the 2DES, the constant density of states condenses into discrete Landau levels (see Fig. 1a). Each Landau level exhibits a Zeeman splitting (omitted in Fig. 1a),  $g\mu_B B$ . The  $g$ -factor of



**Fig. 1.** (a) Schematic picture of the density of states of a 2DES for increasing magnetic field (from bottom to top panels). Gray-shaded areas should indicate occupied states. (b) Experimental PL spectra of a sample with carrier density  $n = 2.3 \times 10^{11} \text{ cm}^{-2}$  at  $T = 2 \text{ K}$  for different magnetic fields

electrons in GaAs,  $g = -0.44$ , arises from band-structure effects ( $\mu_B$  is the Bohr magneton), while additional splittings arise from the strong Coulomb correlations. Each spin-split Landau level has a degeneracy, which is determined by the number of magnetic flux quanta per area,  $eB/h$ . The filling factor  $\nu$  of the system is defined as the number of electrons in the 2DES devided by the number of flux quanta,  $\nu = nh/(eB)$ , where  $n$  is the density of electrons in the 2DES. This means that if  $\nu$  is an integer, we have  $\nu$  completely filled spin-split Landau levels. At these integer values, plateaus are observed in a Hall measurement. Figure 1b displays a series of experimental PL spectra for different magnetic fields of a sample with a carrier density of  $n = 2.3 \times 10^{11} \text{ cm}^{-2}$ . At this relatively large density, the Coulomb interaction between electrons and photoexcited holes is mostly screened and one observes PL lines which can be interpreted as recombination of electrons from the occupied Landau levels in the conduction band with photoexcited holes in the valence band. In the experimental spectra in Fig. 1b, one can see the evolution of occupied Landau levels with magnetic field. At integer  $\nu$  the Fermi energy jumps from one Level to the next lower level, and the PL from the higher level, which is empty now, decreases abruptly. This can especially be seen at  $B = 4.5 \text{ T}$  in Fig. 1b where  $\nu = 2$ . The weak signal from the third Landau level for magnetic fields larger than  $4.5 \text{ T}$  in Fig. 1b stems from photoexcited  $e\text{-}h$  pairs. At the same time, at integer  $\nu$ , there are pronounced anomalies in the intensity and energetic position of the recombination from the lowest Landau level. These can for instance be seen at  $\nu = 2$  ( $B = 4.5 \text{ T}$  in Fig. 1b) and  $\nu = 1$  ( $B = 9 \text{ T}$  in Fig. 1b) where the 2DES is completely spin polarized. There is a variety of literature on these many-particle interaction

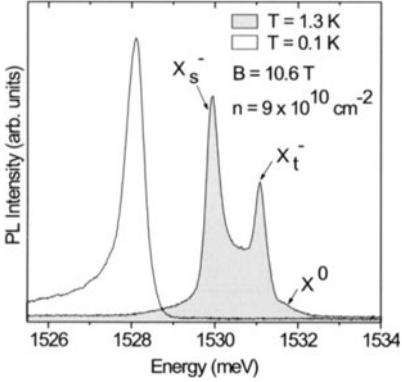


**Fig. 2.** PL spectra for different magnetic fields for a carrier density of  $\sim 0.9 \times 10^{11} \text{ cm}^{-2}$  at a temperature of (a)  $T = 1.8 \text{ K}$  and (b)  $T = 0.1 \text{ K}$

induced effects in PL in high-density samples (see, e.g., [2,5,6]), which shall not be the focus of this work.

### 2.3 Intermediate and Low Electron Density

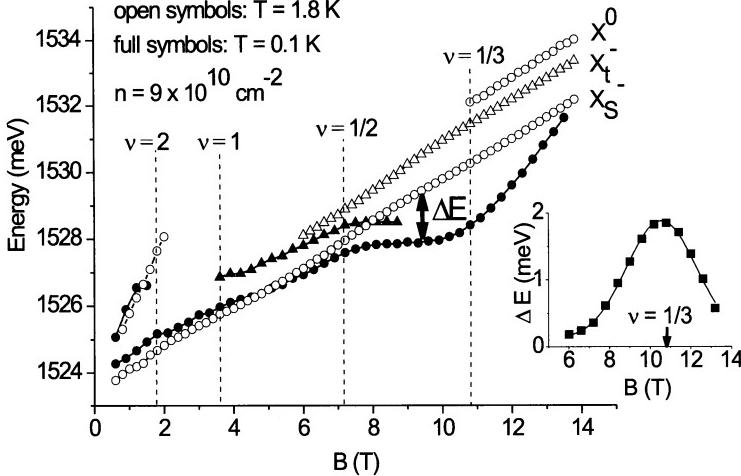
For samples with moderate electron mobility, in the range of  $10^5 \text{ cm}^2/\text{Vs}$ , we expect that for densities below about  $1 \times 10^{11} \text{ cm}^{-2}$  the residual disorder plays a significant role and localization effects of charged excitons might be of importance. In the following we will call this density regime, where we think that in our samples a uniform 2DES just starts to form, the intermediate density regime. Figure 2 shows PL spectra for an electron density of  $0.9 \times 10^{11} \text{ cm}^{-2}$  at a temperature of  $T = 1.8 \text{ K}$  (Fig. 2a) and  $T = 0.1 \text{ K}$  (Fig. 2b). Note that the only difference between the experiments displayed in Figs. 2a and 2b is the different temperature. For filling factors  $\nu > 2$ , i.e., if the lowest Landau level is completely occupied, we find qualitatively the same behavior in both experiments (gray shaded spectra in Figs. 2a and 2b): The observed PL lines show a nearly linear magnetic-field dispersion as observed in the high-density case (Fig. 1a). This can be attributed to recombinations of electrons from the occupied Landau levels with photoexcited holes in the valence band. At  $\nu = 2$ , we find an abrupt crossover to a regime where negatively charged singlet excitons form, which we infer from the nearly quadratic dispersion of the observed PL line for  $\nu < 2$ . A similar behavior, i.e., a crossover between a Landau level-like and a charged exciton regime at  $\nu = 2$ , was reported recently [23,32] for symmetric quantum wells,



**Fig. 3.** Left-circularly polarized photoluminescence spectra for a carrier density of  $9 \times 10^{10} \text{ cm}^{-2}$  and a magnetic field of  $B=10.6 \text{ T}$ , corresponding to  $\nu = 1/3$ , at temperatures  $0.1 \text{ K}$  and  $1.3 \text{ K}$

and was attributed to a breaking of the so called hidden symmetry [11,12] for  $\nu > 2$  [32]. We note that we observe here a similar behavior in *asymmetric* quantum wells.

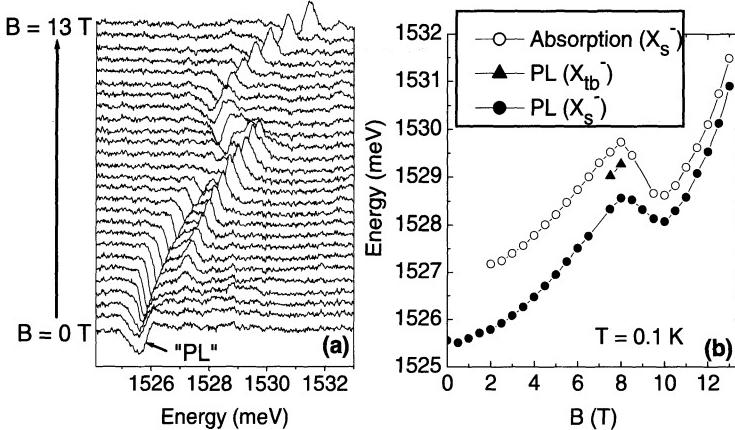
The interesting region, however, where we observe significant differences in the experiments displayed in Fig. 2, is the magnetic field range where  $\nu < 1$ . Here, the measurements at  $T = 1.8 \text{ K}$  (Fig. 2a) exhibit the well known formation of negatively charged excitons: The excitation with lowest energy, the  $X_s^-$ , is visible over the whole magnetic field range for  $\nu < 2$ . For  $\nu < 1$ , the bright triplet exciton  $X_{tb}^-$  [8,33] gradually appears, and, at high fields, the neutral exciton  $X^0$  shows up at the high energy flank of the  $X_{tb}^-$ . The situation changes drastically if we lower the temperature to  $0.1 \text{ K}$  (Fig. 2b). Here, the  $X_{tb}^-$  appears abruptly at  $\nu = 1$ , and, more strikingly, the lowest energy line, presumably the  $X_s^-$ , shows a strong down curvature in energy for filling factors  $\nu < 1/2$ . To emphasize this experimental finding, in Fig. 3 two spectra, taken at  $B = 10.6 \text{ T}$ , which roughly corresponds to filling factor  $\nu = 1/3$ , are compared. Note that, again, the only difference between the two spectra is the different temperature. The spectrum at  $T = 1.3 \text{ K}$  shows the charged and neutral excitons, whereas at  $T = 0.1 \text{ K}$  only a single line, which is strongly redshifted, is observed. To analyze this behavior in more detail, in Fig. 4 the energetic positions of the observed lines at  $T = 1.8 \text{ K}$  (open symbols) and  $T = 0.1 \text{ K}$  (solid symbols) are plotted versus magnetic field for a density of  $0.9 \times 10^{11} \text{ cm}^{-2}$ . One can see that, starting at  $\nu = 1$ , the energies of both,  $X_s^-$  (solid circles in Fig. 4) and  $X_{tb}^-$  (solid uptriangles in Fig. 4), are lowered in the experiment at  $T = 0.1 \text{ K}$  with respect to the experiment at  $T = 1.3 \text{ K}$  (open symbols in Fig. 4). The inset of Fig. 4 shows the experimentally determined anomaly  $\Delta E$  versus magnetic field. The most striking result here is that a thermal energy of  $\sim 2 \text{ K}$  ( $\sim 0.2 \text{ meV}$ ) is sufficient to completely destroy the anomaly which has a strength of about  $2 \text{ meV}$ . This rules out any trivial localization effect of excitons, since then one would expect that at least a thermal energy of also about  $2 \text{ meV}$  would be necessary to delocalize the excitons.



**Fig. 4.** Experimentally observed mode positions at  $T = 1.8$  K (open symbols) and  $T = 0.1$  K (solid symbols). The inset shows the anomaly  $\Delta E$  versus magnetic field

To get more information about the possible origin of the anomaly, we have also performed direct absorption measurements at low temperatures. Absorption experiments give direct information about the oscillator strengths of the observed excitations, while the intensities of PL lines can also be dominated by effects like localization of excitons at impurities or potential fluctuations. Figure 5a shows a series of absorption spectra using white light for different magnetic fields. The displayed spectra  $S(E)$  are calculated from the measured transmission spectra  $T(E)$  by the relation  $S(E) = -\ln(T(E)/T_0(E))$ , where  $T_0(E)$  is the measured transmission without sample. Due to the relatively strong illumination with white light, also PL recombinations are initiated and add to the transmission signals. Therefore, in the normalized spectra in Fig. 5a, PL lines appear as dips, while absorption lines appear as peaks. Figure 5b shows the extracted dispersions of the PL and the absorption lines which we attribute to the  $X_s^-$  and  $X_{tb}^-$  excitons. The interesting result is that the absorption shows the same anomaly as the PL line, which is a strong indication that the anomaly is an intrinsic effect and is not caused by localization of excitons. In the case of localization, one would expect that the absorption should still reflect dominantly the intrinsic structure of the system, whereas the PL might be dominated by extrinsic effects, like exciton localization.

Interestingly, if we lower the density to the range  $\sim 10^{10} \text{ cm}^{-2}$ , the anomaly disappears completely, even in experiments at very low temperature. This can be seen in Fig. 6, where contour plots of PL spectra are shown for three different densities. One can see the well-known behavior of the charged excitons, and, for filling factors  $\nu < 1/3$ , the so called dark triplet exciton  $X_{td}^-$ , which has been observed in experiments at very low temperature [7,8]. This charged triplet exciton was found to exhibit the singlet-triplet crossing



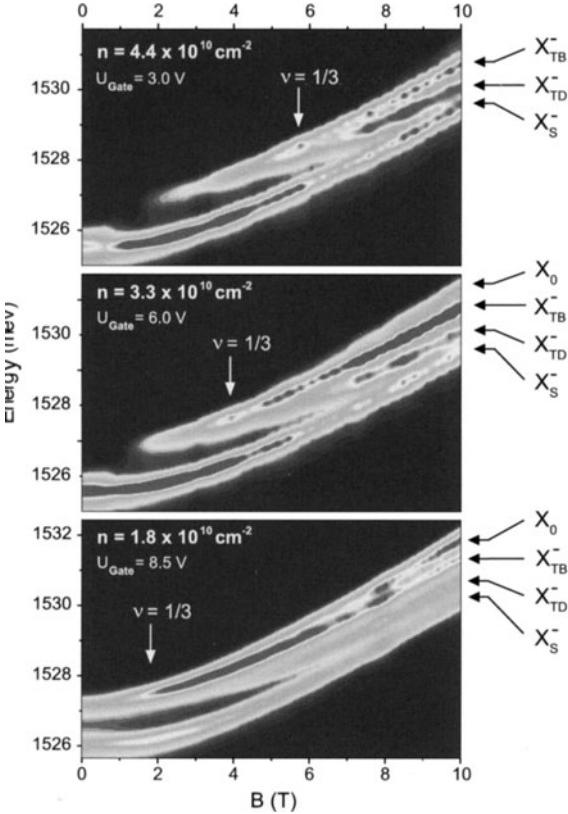
**Fig. 5.** (a) Normalized (see text) transmission spectra (unpolarized) for different magnetic fields [33,34]. (b) Energetic positions of PL and absorption lines versus magnetic field

at high magnetic fields [33,34]. However, the important information of Fig. 6 for the investigation here is that in the range of filling factor  $\nu = 1/3$  no anomalous dispersion of the charged excitons can be observed.

Before we give an explanation for the anomaly around  $\nu = 1/3$ , we want to summarize at this point the relevant facts from the experiments: (a) The anomaly is not seen at *higher* electron densities where no charged excitons but usual electrons exist, and is also not seen for *lower* electron densities where exclusively charged excitons are present [8]. Also for higher mobility samples ( $\sim 5 \times 10^6 \text{ cm}^2/\text{Vs}$ , not shown here) it is not present. Because of the low mobility and relatively low density of the sample, excitons are expected to remain localized. (b) The anomaly appears near  $1/3$ , i.e., excitons are near an incompressible liquid. (c) The most intriguing observation is that *a very small thermal energy* ( $\ll 2 \text{ meV}$ ) *is required to destroy the anomaly*. (d) The anomaly does not appear near  $\nu = 1, 2$  and is therefore an indication that the lowest-energy charged excitations, the quasiholes (for reasons to be discussed below) are perhaps involved in the process. The quasielectrons are predicted to have higher energies [28].

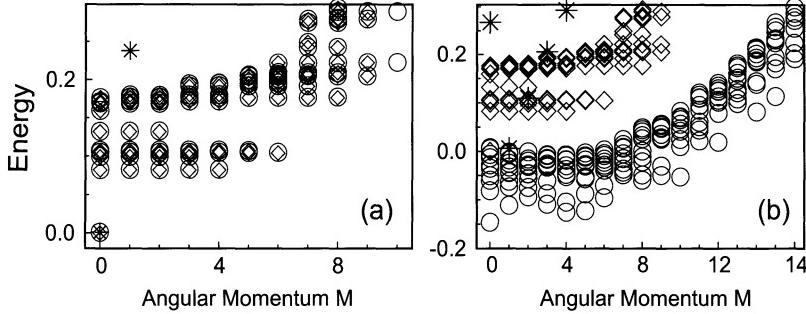
### 3 Theoretical Model: Creation of Fractionally-Charged Quasiholes

In our explanation of the observed anomaly [10] we assume that, as a result of potential fluctuations due to impurities in the system, excitons remain localized but they are in close proximity to the incompressible liquid at  $\nu = 1/3$ . Two of us recently investigated a system [35] where a parabolic quantum dot (QD) [36] is coupled (via the Coulomb force) to a 2DES which is in a



**Fig. 6.** Contour plots of PL experiments at  $T = 0.1 \text{ K}$  for different carrier densities

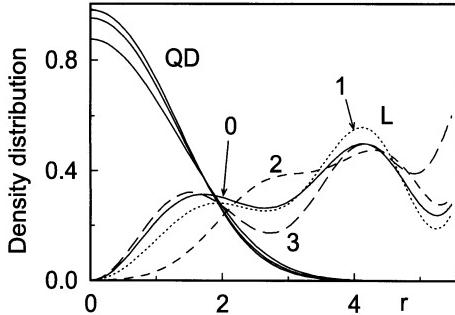
$\nu = 1/3$  Laughlin state. Electrons in the dot are confined by a parabolic potential [36],  $V_{\text{conf}}(x, y) = \frac{1}{2}m^*\omega_0^2(x^2 + y^2)$ , where  $\omega_0$  is the confinement potential strength and the corresponding oscillator length is  $l_{\text{dot}} = (\hbar/m^*\omega_0)^{1/2}$ . Calculating the low-energy excitations of that quantum dot-liquid system (named a *qd-liquid*) we found that for a single electron in the dot the physics is somewhat similar to that of a point impurity in a  $\nu = \frac{1}{3}$  liquid state investigated earlier [37]. In this case, the QD emits a fractionally-charged *quasihole* ( $e/3$ ) that orbits around the QD, as evidenced from the charge-density calculations [35,37]. Here we propose that the observed anomaly is related to the qd-liquid where the QD contains a charged exciton. The QD in our model of Ref. [35] represents a localized exciton (charged or neutral) in the present case, and perturbs the incompressible fluid due to its close proximity by creating fractionally-charged defects. Details on the formal aspects of our theory can be found in Ref. [35]. We model the incompressible state at  $\nu = 1/3$  filling using the spherical geometry [29] for six electrons. Electrons are treated as spinless particles corresponding to the state described by the Laughlin wavefunction [25]. We consider the QD size  $l_{\text{dot}} = 15 \text{ nm}$  and the



**Fig. 7.** Energy (in units of Coulomb energy) versus the azimuthal rotational quantum number  $M$  for an isolated quantum dot (\*), a two-dimensional electron liquid ( $\diamond$ ), and a qd-liquid ( $\circ$ ). The QD of the qd-liquid either contains  $(1e, h)$  [in (a)], or  $(2e, h)$  [in (b)]

liquid-dot separation  $d = 1.5l_0$ . The QD contains either a pair of electron and hole  $(e, h)$  (charge-neutral QD), or  $(2e, h)$  (charged QD). Figure 7 shows the energy spectra for the qd-liquid where the QD contains either  $(e, h)$  [in (a)] or  $(2e, h)$  [in (b)]. In the figures, the energy spectra of isolated dots (\*), an incompressible liquid at  $\nu = 1/3$  state ( $\diamond$ ) and the binding energy of the QD to the incompressible liquid ( $\circ$ ) are plotted for comparison. From Fig. 7a we can see that for a charge-neutral dot there is no dispersion of the energy as a function of  $M$ , and most importantly, the incompressible liquid is not influenced by the dot at all. On the other hand, the energy of the qd-liquid is significantly lowered for a charged QD, as compared to the isolated QD or the incompressible liquid without the dot [Fig. 7b]. This is in line with the experimental observation where only the charged excitons show the anomaly by lowering the energy.

Figure 8 exhibits the electron density distribution in the liquid (L) and in the dot (QD) for the lowest states and for a given angular momentum of the qd-liquid system. The electron (or hole) system in the dot is close to the ground state of an isolated dot, i.e., the influence of the incompressible liquid on it is very small. On the other hand, the low-lying excited states of the qd-liquid can be described by the ionization process [35,37] as emission of a quasi-hole: Since the net charge of a QD is negative, the ground state of the qd-liquid can be considered as a QD plus three quasiholes. If we increase the angular momentum of the 2D electrons, one of the quasiholes moves away from the QD. This is inferred from the calculated charge distribution of electrons around the QD, where a local minimum corresponds to the quasi-hole moving away from the QD as the angular momentum is increased. The position of the local minimum at different angular momenta of the charge density correspond to the orbit radius of the quasi-hole [35,37]. We have evaluated the quasi-hole creation energy in a qd-liquid. For the Laughlin state in a pure 2DES it is  $0.0276 e^2/\epsilon l_0$  [26]. For the qd-liquid, the corresponding



**Fig. 8.** Charge-density profile of the ground state and low-lying excitations of electrons in the dot (QD) and the liquid (L) of a qd-liquid at  $\nu = 1/3$  (for  $M = 0 - 3$ , as indicated in the figure) corresponding to Fig. 7b

value is 0.32 meV and is expected to decrease a little further with increasing number of electrons in the system representing the incompressible liquid [29]. This result indicates that the small thermal energy of about 0.2 meV required to destroy the anomaly is in fact, the quasi-hole energy gap.

## 4 Conclusion

PL experiments on a 2DES subjected to a quantizing magnetic field exhibit a significant lowering of exciton energies at and around  $\nu = 1/3$ . This anomalous dispersion is explained as due to the perturbation of the incompressible liquid at  $\nu = 1/3$  by localized charged excitons which results in the creation of fractionally-charged quasiholes in the liquid.

## References

1. D. C. Tsui et al., Phys. Rev. Lett. **48**, 1559 (1982).
2. A. J. Turberfield et al., Phys. Rev. Lett. **65**, 637 (1990).
3. I. V. Kukushkin et al., Phys. Rev. Lett. **82**, 3665 (1999).
4. I. V. Kukushkin et al., Phys. Rev. Lett. **85**, 3688 (2000).
5. B. B. Goldberg et al., Phys. Rev. Lett. **65**, 641 (1990).
6. L. Gravier et al., Phys. Rev. Lett. **80**, 3344 (1998)
7. G. Yusa et al., Phys. Rev. Lett. **87**, 216402 (2001).
8. C. Schüller et al., Phys. Rev. B **65**, 081301(R) (2002).
9. K.-B. Broocks et al., Phys. Rev. B **66**, 041309(R) (2002).
10. C. Schüller et al., Phys. Rev. Lett. **91**, 116403 (2003).
11. A. H. MacDonald and E. H. Rezayi, Phys. Rev. B **42**, 3224 (1990).
12. A. B. Dzyubenko and Yu. E. Lozovik, J. Phys. A **24**, 415 (1991).
13. V. M. Apalkov and E. I. Rashba, Phys. Rev. B **46**, 1628 (1992).
14. G. Finkelstein et al., Phys. Rev. Lett. **74**, 976 (1995).
15. A. J. Shields et al., Phys. Rev. B **52**, R5523 (1995).
16. G. Finkelstein et al., Phys. Rev. B **53**, R1709 (1996).
17. A. Wojs and P. Hawrylak, Phys. Rev. B **51**, 10880 (1995).
18. J. J. Palacios et al., Phys. Rev. B **54**, R2296 (1996).
19. D. M. Whittaker and A. J. Shields, Phys. Rev. B **56**, 15185 (1997).

20. Y. Kim, F. M. Munteanu et al., Phys. Rev. B **61**, 4492 (2000).
21. M. Hayne et al., Phys. Rev. B **59**, 2927 (1999).
22. C. Riva et al., Phys. Rev. B **63**, 115302 (2001).
23. D. Gekhtman et al., Phys. Rev. B **54**, 10320 (1996).
24. D. Brinkmann et al., Phys. Rev. B **60**, 4474 (1999); P. Gilliot et al., Phys. Rev. B **60**, 5797 (1999).
25. R.B. Laughlin, Phys. Rev. Lett. **50**, 1395 (1983); Surf. Sci. **142**, 163 (1984).
26. T. Chakraborty, Phys. Rev. B **31**, 4026 (1985).
27. See, for example, G.S. Boebinger et al., Phys. Rev. Lett. **55**, 1606 (1985); R.G. Clark et al., *ibid.* **62**, 1536 (1989); J.P. Eisenstein et al., *ibid.* **62**, 1540 (1989).
28. T. Chakraborty and P. Pietiläinen, *The Quantum Hall Effects* (Springer, 1995), 2nd Edition; T. Chakraborty, Adv. Phys. **49**, 959 (2000); T. Chakraborty, in *Handbook on Semiconductors*, vol. 1, edited by P.T. Landsberg (Elsevier, New York, 1992), Ch. 17.
29. F.D.M. Haldane, Phys. Rev. Lett. **51**, 605 (1983); G. Fano et al., Phys. Rev. B **34**, 2670 (1989); R. Morf and B.I. Halperin, Phys. Rev. B **33**, 2221 (1986).
30. R. de-Picciotto et al., Nature (London) **389**, 162 (1997); L. Saminadayar et al., Phys. Rev. Lett. **79**, 2526 (1997).
31. J. J. LePore, J. Appl. Phys. **51**, 6441 (1980).
32. E.I. Rashba and M.D. Sturge, Phys. Rev. B **63**, 045305 (2000).
33. A. Wojs et al., Phys. Rev. B **62**, 4630 (2000).
34. T. Vanhoucke et al., Phys. Rev. B **65**, 233305 (2002).
35. V.M. Apalkov and T. Chakraborty, Physica E **14**, 289 (2002).
36. P.A. Maksym and T. Chakraborty, Phys. Rev. Lett. **65**, 108 (1990); T. Chakraborty, Comments Condens. Matter Phys. **16**, 35 (1992); T. Chakraborty, *Quantum Dots* (North-Holland, Amsterdam, 1999); T. Chakraborty, F. Peeters, and U. Sivan (Eds.), *Nano-Physics & Bio-Electronics: A New Odyssey* (Elsevier, Amsterdam, 2002).
37. E.H. Rezayi and F.D.M. Haldane, Phys. Rev. B **32**, 6924 (1985).

# Three-Dimensional Lithography of Photonic Crystals

A. Blanco<sup>3,4</sup>, K. Busch<sup>2,4</sup>, M. Deubel<sup>1,4</sup>, C. Enkrich<sup>1,4</sup>,  
G. von Freymann<sup>3,4,6</sup>, M. Hermatschweiler<sup>1,4</sup>, W. Koch<sup>1,4</sup>, S. Linden<sup>3,4</sup>,  
D. C. Meisel<sup>3,4</sup>, G. A. Ozin<sup>6</sup>, S. Pereira<sup>2</sup>, C. M. Soukoulis<sup>5</sup>, N. Tétreault<sup>6</sup>,  
and M. Wegener<sup>1,3,4</sup>

<sup>1</sup> Institut für Angewandte Physik, Universität Karlsruhe (TH)  
Wolfgang-Gaede-Straße 1, 76131 Karlsruhe, Germany

<sup>2</sup> Institut für Theorie der Kondensierten Materie, Universität Karlsruhe (TH)  
Wolfgang-Gaede-Straße 1, 76131 Karlsruhe, Germany

<sup>3</sup> Institut für Nanotechnologie  
Forschungszentrum Karlsruhe in der Helmholtz-Gemeinschaft  
76021 Karlsruhe, Germany

<sup>4</sup> DFG-Center for Functional Nanostructures (CFN)

<sup>5</sup> Ames Laboratory and Department of Physics and Astronomy  
Iowa State University, Ames, Iowa 50011, U.S.A.

<sup>6</sup> Department of Chemistry, University of Toronto  
80 St. George Street, Toronto, Ontario, M5S 3H6, Canada

**Abstract.** We present our work on the fabrication of three-dimensional Photonic Crystal templates by holographic lithography and direct laser writing. Both optical methods are highly qualified for the production of top-quality polymeric three-dimensional Photonic Crystals and allow in combination the controlled incorporation of defects and waveguides in large-area structures. Scanning electron micrographs and optical spectra of different Photonic Crystals are presented and compared with numerical simulations. Our roadmap to infiltrate the polymeric templates with high refractive-index materials is outlined.

## 1 Introduction

Photonic Crystals (PhCs) are optical materials which exhibit a periodic variation of the refractive index in one, two or three dimensions. The properties of electromagnetic modes in these PhCs are determined by multiple scattering and can be described by a band structure [1,2]. For certain three-dimensional (3D) PhCs, Bragg scattering leads to the formation of a complete photonic band gap (PBG), i.e. a frequency range in which the propagation of electromagnetic waves is forbidden for any wave vector and polarization. Intentionally introduced defects in the crystal structure provide the opportunity of "molding the flow of light" [3]. These defects give rise to localized photonic states inside the PBG and can be used as functional elements for photonic devices like microcavities or waveguides. Compared to their conventional counterparts, these elements can be realized on a much smaller scale. Ultimately, the goal is to create integrated photonic circuits based on PhCs.

While the theoretical description of light in PhCs is very advanced, insufficient quality of 3D PhCs is still a major issue on the experimental side. Many different fabrication methods have been proposed to overcome this problem. Each of these methods has its own advantages and disadvantages. For example, self-assembly allows the fabrication of large-area crystals [4] but is very sensitive to unintentional defects. On the other side, layer-by-layer techniques based on wafer bonding [5] yield high-quality samples but are extremely complex and require expensive equipment.

Our approach to 3D PhC fabrication comprises two compatible and complementary lithographic methods: Holographic lithography (HL) and direct laser writing (DLW). These optical methods, which are highly qualified for the production of top-quality polymeric 3D PhCs, are described in detail in sections 2 and 3.

The refractive index of the used polymer is only 1.6. Consequently, a supplementary inversion process of the polymeric PhCs with a high refractive index material is necessary to open up a complete PBG. Our roadmap to accomplish this task is outlined in section 4.

## 2 Holographic Lithography

In order to fabricate matter distributions on a sub-micron scale one can make use of multiple-beam interference patterns. Such light distributions can be periodic or quasi-periodic with a period on the order of half the wavelength. They are interesting for the fabrication of PhCs since a corresponding matter distribution can be obtained by exposing a photoresist to the interference pattern. The resulting porous material can serve as a template for PhCs [6,7,8,9].

The crystallographic symmetries of the so-called “umbrella-like” beam configuration are discussed in some detail. We for the first time give experimental parameters for this geometry that allow to fabricate PhC templates that after silicon inversion are predicted to have a full PBG at near-infrared frequencies [10].

### 2.1 The Photoresist SU-8

We use the commercially available photoresist SU-8 from MicroChem. It consists of the epoxy EPON SU-8 and a photoinitiator both dissolved in gamma-butyrolactone (GBL). Upon irradiation by near-UV light (350–400 nm) the photoinitiator generates an acid. The spatial acid concentration is an image of the irradiation dose. In a post-exposure bake the latent image is converted into a cross-linking density by cationic polymerization during this thermal treatment.

The cross-linking degree determines the solubility in the “developing” solvent. GBL or another appropriate solvent is used in this step. Thus, sufficiently illuminated resin remains (“negative” photoresist) whereas underexposed resin is washed away. The remaining material therefore has a shape

that follows the surface of the threshold irradiation which is the boundary between over- and under-exposed regions. Effectively, the interference pattern is stored in a digitized form: It is a porous air-polymer-structure showing the iso-dose surface for the threshold-dose value.

## 2.2 The Crystallography of Multiple-beam Interference Patterns

For band-structure calculations it is of importance to know the structure that is obtained for a given beam configuration. The shape of the resulting template is given by the iso-dose surfaces.

The  $N$  interfering light beams are treated as plane waves with the electric field vector as real part of a complex quantity:

$$\mathbf{E}_n(\mathbf{r}, t) = \text{Re}\{\mathbf{E}_n^0 e^{i(\mathbf{k}_n \cdot \mathbf{r} - \omega t)}\} \quad \text{with integer } n = 1 \dots N. \quad (1)$$

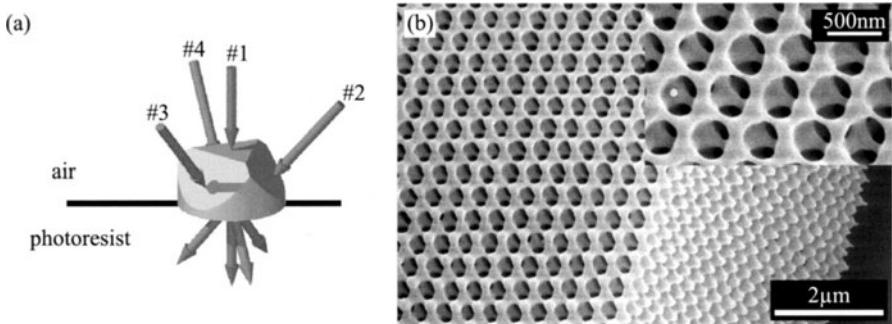
All  $|\mathbf{k}_n|$  are the same since the beams are obtained by beam splitting. The exposure dose  $\mathcal{D}(\mathbf{r})$  is a Fourier sum [6,7]:

$$\mathcal{D}(\mathbf{r}) \propto \left| \sum_{n=1}^N \mathbf{E}_n^0 e^{i(\mathbf{k}_n \cdot \mathbf{r} - \omega t)} \right|^2 = \sum_{n,m=1}^N a_{nm} e^{i\mathbf{G}_{nm} \cdot \mathbf{r}}. \quad (2)$$

The “form factors”  $a_{nm} := \mathbf{E}_n^0 \cdot \mathbf{E}_m^{0*}$  are obtained from the relative amplitudes, polarizations and phases of the beams. The differences of the wave vectors  $\mathbf{G}_{nm} := \mathbf{k}_n - \mathbf{k}_m$  are reciprocal lattice vectors. The translational symmetry of the interference pattern is determined by the  $\mathbf{G}_{nm}$ , whereas the shape of the motif results from the form factors  $a_{nm}$ . Lattice and motif in general have different point symmetries, resulting in some overall crystal symmetry common to both.

In order to obtain 3D structures one has to apply at least four beams. At present there are mainly two four-beam configurations under discussion. The “two-planes” geometry has two pairs of beams traveling in planes perpendicular to each other. In this geometry the substrate has to be passed by two beams; since beams are counter-propagating, rather small lattice constants result. We therefore prefer the experimentally advantageous “umbrella-like” configuration: Three beams are equally distributed on a cone with apex angle  $\gamma$ . The fourth beam is directed along the axis of the cone. All beams are pointing in the same half-space for  $\gamma < 90^\circ$ .

In this symmetric configuration the translational symmetry in general is rhombohedral with simple cubic, face-centered cubic (fcc) or body-centered cubic as special cases. It is assumed that fcc translational symmetry is beneficial for a complete band gap [1]. This requires  $\gamma = \gamma_{\text{fcc}} \approx 38.94^\circ$  which in turn can be realized with an immersion coupling prism only [9]. Fig. 1(a) shows the arrangement of the laser beams. The resulting structures (see scanning electron microscope (SEM) image Fig. 1(b)) have fcc translational symmetry. For this apex angle  $\gamma_{\text{fcc}}$  the highest possible point symmetry of the motif is



**Fig. 1.** (a) Four laser beams arranged in the “umbrella-like” geometry and immersion coupling prism. (b) SEM image of a resulting structure with fcc translational symmetry [9]

rhombohedral. An example of a PhC with fcc translational symmetry but rhombohedral overall symmetry is the “Yablonovite” structure the complete bandgap of which has been verified experimentally in the radio-frequency regime [11]. Therefore we search for conditions the  $a_{nm}$  have to meet such that the overall symmetry is rhombohedral. For this purpose we compare the dose (2) with the geometrical structure factors of the rhombohedral space groups [10]. We also call for inversion symmetry which is believed to be advantageous for a band gap. Linearly polarized light meets this demand. Still many solutions for the  $E_n^0$  are possible. We use the remaining freedom to maximize the interference contrast [10]. Fig. 2 shows the rhombohedral unit cell with its motif as resulting from the above considerations. The above symmetry considerations for the motif were not yet taken into account in our early experiments (Fig. 1(b)).

Band-structure calculations show a complete bandgap after silicon inversion of the resin template. After removing the template 37% of the volume is silicon. The gap/midgap ratio of the fundamental gap is about 5% [10].

### 3 Direct Laser Writing

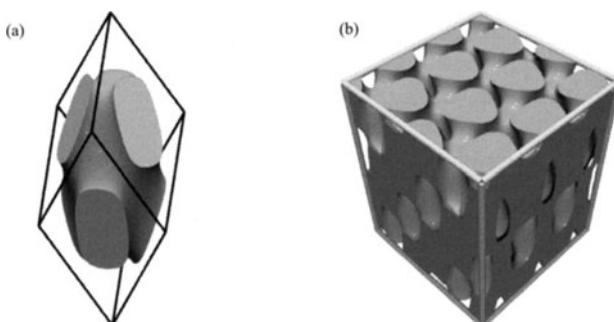
Holographic lithography is capable of producing samples with defect-free areas of several square millimeters and offers a high flexibility in tailoring the interior of the unit cell. However, due to the fabrication principle, these structures are strictly periodic, i.e., defect structures can not be included in a controlled fashion. Therefore, a second complementary technique is required to inscribe functional elements into bulk PhCs provided by HL. An excellent candidate for this task is DLW through multi-photon polymerization (MPP) [12,13,14].

### 3.1 Multi-photon Polymerization

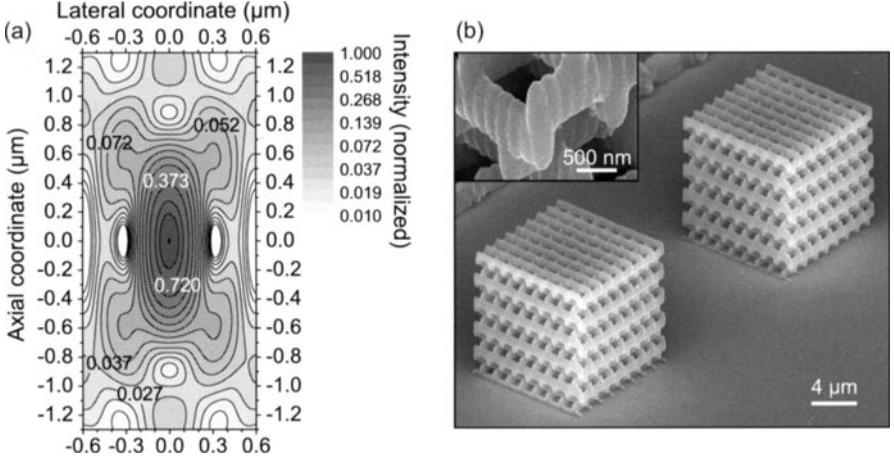
In MPP DLW, a photoresist exhibiting an intensity threshold for exposure is illuminated by laser light whose photon energy is insufficient to expose the photoresist by a one-photon absorption process. If this laser light is tightly focused into the resist, however, the light intensity inside a small volume element (“voxel”) inclosing the focus may become sufficiently high to exceed the exposure threshold by multi-photon processes. By scanning the focus relatively to the photoresist, in principle, any 3D structure consisting of these voxels may be written directly into the photoresist. The size and shape of the exposed voxels depend on the isophotes of the microscope objective and the multi-photon exposure threshold of the photosensitive medium. The calculated isophotes of the microscope objective that is used in our experiments are shown in Fig. 3(a). The isophotes in vicinity of the geometrical focus exhibit a typical near-ellipsoidal shape with an aspect ratio of about 2.7.

### 3.2 Experimental Realization

To guarantee compatibility with HL, we use the SU-8 negative photoresist in the DLW fabrication process (see section 2.1). DLW then simply may be added as a second exposure step to the holographic fabrication process prior to development of the PhC. Additionally, the cationic polymerization mechanism of SU-8 is advantageous, since the monomers do not polymerize directly after exposure, leading to negligible refractive-index contrast of exposed and unexposed photoresist. Therefore, structures that have already been exposed will not affect additional exposure steps. A further advantage is that SU-8 is solid during the writing process. This does not only make the sample handling much more comfortable, but also offers a higher degree of freedom in the scanning pattern, since successively written structures do not have to be interconnected immediately.



**Fig. 2.** Computed image of the photoresist template. (a) The rhombohedral unit cell. (b) Part of the structure [10]



**Fig. 3.** (a) Calculated isophotes of the NA=1.4 microscope objective used in our experiments. The intensity is normalized to the value in the origin (geometrical focus). (b) SEM image of two layer-by-layer structures with  $a = 1.5 \mu\text{m}$ . In the inset, the single voxels that build the rods can be distinguished. The surface roughness of the voxels is determined by the resolution of the resist itself

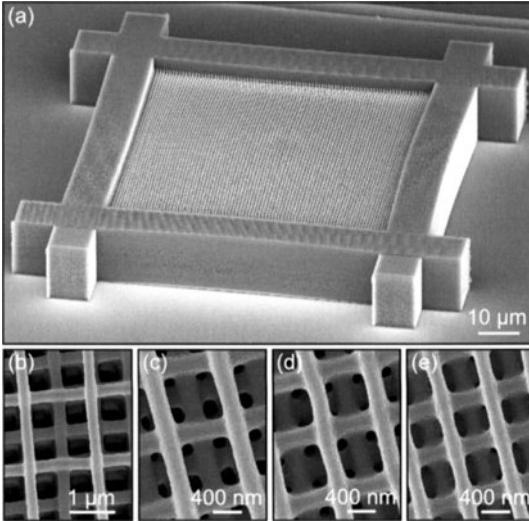
Prior to exposure, a  $20 \mu\text{m}$  thick SU-8 photoresist film is spin-coated on a microscope cover slide and solidified by a soft-bake process. The laser, a regeneratively amplified Ti:sapphire laser system (Spectra Physics Hurricane, pulse duration 120 fs) is tuned to a central wavelength of 800 nm, where the one-photon absorption of SU-8 is negligible [15]. The output beam is attenuated by a halfwave-plate/polarizer combination and after beam expansion, typically a few tens of nanojoules of single pulse energy are focused into the photoresist by the  $100\times$  oil-immersion microscope objective (numerical aperture NA=1.5) of a custom designed inverted microscope (Leica). This photoresist sample is placed on a three-axis piezoelectric scanning stage (Physik Instrumente) that provides a resolution of 5 nm at a full scanning range of  $200 \mu\text{m} \times 200 \mu\text{m} \times 20 \mu\text{m}$ . A personal computer controls the scanning operation of the piezo stage and synchronizes its movement with the pulsing of the laser system via a laser controlling interface. After DLW of a preprogrammed pattern, the exposed sample is postbaked and developed as described in section 2 for the holographic PhC templates.

### 3.3 Direct Laser Writing of Three-dimensional Photonic Crystals

#### 3.3.1 Woodpile Photonic Crystals

Owing to their simple geometrical structure, the so called layer-by-layer [16] or woodpile [17] PhCs are ideal structures for DLW.

Fig. 3 (b) shows two samples of this geometry fabricated with our setup:

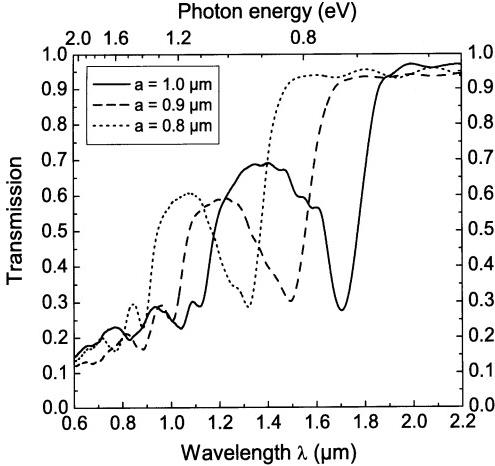


**Fig. 4.** (a) Larger layer-by-layer structure with a massive wall that prevents bending and reduces lattice distortions due to polymer shrinkage during polymerization. (b)-(e) Closeups of samples with  $a = 1.5 \mu\text{m}$ ,  $a = 1.0 \mu\text{m}$ ,  $a = 0.8 \mu\text{m}$  and  $a = 0.65 \mu\text{m}$ , displaying the high fabrication accuracy

Layers of straight parallel rods with a center-to-center distance  $a$  are stacked to a 3D lattice. Adjacent layers have the orientation of the rods rotated by  $90^\circ$  and second nearest neighboring layers are shifted by a distance of  $a/2$  perpendicular to the rod axes. This stacking sequence repeats itself after every four layers with a lattice constant  $c$ . For  $c/a = \sqrt{2}$ , this 3D lattice exhibits a fcc unit cell with a two-rod basis and can be derived from a diamond lattice by replacing the (110) chains of lattice points with rods [16]. With our fabrication technique, these rods are built up of joining individual voxels that are exposed by a single laser pulse each. In the inset of Fig. 3 (b) the voxels are clearly distinguishable and exhibit the typical near-ellipsoidal shape, as expect from the calculated isophotes.

In order to do optical experiments on such samples, the area of the structures needs to be increased. Fig. 4 (a) shows a sample consisting of 40 layers with an area of  $70 \mu\text{m} \times 70 \mu\text{m}$ . The in-plane rod distance  $a$  is  $0.8 \mu\text{m}$  and  $c/a$  is adjusted to  $\sqrt{2}$  for fcc symmetry. To reduce lattice distortions and sample bending due to polymer shrinkage, a massive wall is polymerized around all PhC structures by the same technique. Although this is a quite large structure for DLW with that resolution, the writing time of the PhC lattice is just 25 minutes with a scanning algorithm optimized for the layer-by-layer structures.

With the highly flexible DLW technique, it is possible to write a wide variety of lattice constants and filling fractions by changing the scanning pattern and the laser pulse energy, respectively. Figs. 4 (b)-(e) show an overview over several structures that we have fabricated. All lattices were designed with  $c/a = \sqrt{2}$  for fcc symmetry. The in-plane rod distance  $a$  here ranges from rather large  $1.5 \mu\text{m}$  down to  $650 \text{ nm}$ .



**Fig. 5.** Transmission spectra of three different layer-by-layer samples with in-plane rod spacings of  $a = 1.0 \mu\text{m}$  (solid line),  $a = 0.9 \mu\text{m}$  (dashed line), and  $a = 0.8 \mu\text{m}$  (dotted line)

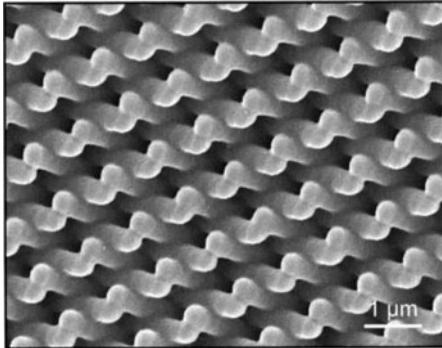
To move the fundamental PBGs to wavelengths in the telecommunication regime of  $1.3 \mu\text{m}$  and  $1.5 \mu\text{m}$ , it is necessary to have an in-plane rod spacing on the order of  $1 \mu\text{m}$  and less. Therefore, we prepared samples with rod spacings of  $1.0 \mu\text{m}$ ,  $0.9 \mu\text{m}$ , and  $0.8 \mu\text{m}$ , all samples with 24 layers and  $c/a = \sqrt{2}$ . The sample area of  $100 \mu\text{m} \times 100 \mu\text{m}$  is sufficiently large for optical spectroscopy as well as for potential devices. Transmission and reflection of these structures are measured with a Fourier-transform infrared (FTIR) spectrometer (Bruker Equinox 55, NIR halogen source) combined with an infrared microscope (Bruker Hyperion 2000,  $36 \times$  Cassé-grain objectives, NA=0.5, liquid N<sub>2</sub>-cooled InSb detector). The samples are aligned with their surfaces perpendicular to the optical axis, which corresponds to the  $\Gamma$ -X direction for fcc symmetry. A circular area with  $50 \mu\text{m}$  diameter is defined by an aperture in the light path of the microscope. The transmission and reflection spectra are normalized to the bare cover slide and a silver mirror, respectively.

Figure 5 (a) shows the measured transmission spectra for the three different samples. The fundamental stop bands are clearly visible as pronounced dips in transmission and shift from  $1.7 \mu\text{m}$  for the sample with  $a = 1.0 \mu\text{m}$ , to  $1.5 \mu\text{m}$  for the sample with  $a = 0.9 \mu\text{m}$ , and to  $1.3 \mu\text{m}$  for the sample with  $a = 0.8 \mu\text{m}$ . Higher-order stop bands are also observed.

To check the quality of our structures, we have also compared these transmission and reflection measurements with spectra that were calculated using an S-matrix technique. The overall agreement between simulations and measurements is very good and almost quantitative [18].

### 3.3.2 SP<sub>2</sub> Photonic Crystals

Related to the woodpile structure is the so-called "SP<sub>2</sub>" structure from the family of tetragonal Slanted Pore PhCs, which has been proposed by O. Toader only very recently [19]. For special parameters this architecture is



**Fig. 6.** Slanted Pore PhC of the  $S/[1, 1] \oplus [-1, -1]^{(0.5,0)}$  family with  $a = 1.0 \mu\text{m}$ ,  $c = 1.41 \mu\text{m}$

indeed equivalent to the woodpile structure but it suggests new approaches and opportunities for fabrication. Importantly, it is extremely robust against fabrication tolerances. From the SP<sub>2</sub> PhC architecture described in [19], we chose the  $S/[1, 1] \oplus [-1, -1]^{(0.5,0)}$  structure for fabrication by DLW. It has a tetragonal unit cell with  $a_1 = a_2 = a = 1.0 \mu\text{m}$  and  $c = 1.41 \mu\text{m}$ . The rod diameter is adjusted to about 360 nm so that rods of adjacent layers touch to give the structure mechanical stability. Within the rods, the voxels are now arranged with their axes tilted by 45° with respect to the rod axis, resulting in a near-elliptical rod cross section with an aspect ratio of only about 1.9 compared to 2.7 of the woodpile geometry. Figure 6 (a) shows a SEM image of a sample with an area of  $40 \mu\text{m} \times 40 \mu\text{m}$ , which is large enough for optical experiments as well as for potential devices. As in the case of the woodpile-type samples, we have measured transmission and reflection spectra of these samples (not shown) and find a pronounced dip in transmission and a peak in reflection around  $1.8 \mu\text{m}$  wavelength, indicating the fundamental stop band. Higher order features around  $0.9 \mu\text{m}$  are also visible, again underlining the high optical quality of our samples.

## 4 Outlook: Inversion

The logical next step and challenge is the infiltration of our 3D PhC templates with dielectrics which exhibit a large refractive index and/or with metals. Currently, two methods are investigated by us, namely chemical vapor deposition (CVD) of silicon and electrochemical deposition of silver.

Silicon CVD is a standard technique, which is commonly used for the fabrication of inverted opals [20]. However, the temperature stability of SU-8 inhibits a direct inversion of our templates. Therefore, the templates are first infiltrated with Silica ( $\text{SiO}_2$ ) in a process which can be accomplished at room temperature. In order to remove the SU-8, a new substrate is placed on top of the structure and the old substrate is removed. Next, the SU-8 resist is burned by heating the sample up to 500°C for several hours. The

resulting SiO<sub>2</sub> inverted template is compatible with the silicon CVD process. Here, the sample is heated up to 375°C at an ambient pressure of 10<sup>-6</sup> mbar. Disilane (Si<sub>2</sub>H<sub>6</sub>) gas is added to the reaction chamber which decomposes on the surface of the inverted template due to the high temperature. This leads to the formation of a silicon layer whose thickness is monitored by a sensing crystal. After the silicon CVD process, the SiO<sub>2</sub> is removed with an aqueous solution of HF.

Electrodeposition is an alternative and a versatile technique suitable for the infiltration of our SU-8 PhC templates with metals or other materials. For this purpose, we are currently developing a computer-controlled electrodeposition setup which comprises a standard three-electrode configuration. Clearly, a conductive substrate is required for the electrodeposition process. Unfortunately, coating the substrate with a thin metal film is not compatible with the holographic fabrication process of the SU-8 templates since the metal coating gives rise to unwanted reflections. A good alternative is the use of indium-tin-oxide (ITO). We have deposited ITO layers which are simultaneously almost transparent and exhibit a sufficient conductance for the electrodeposition process. The deposition of silver in the voids of the SU-8 templates is currently under way. The first results are very encouraging.

## References

1. E. Yablonovitch, Phys. Rev. Lett. **58**, 2059 (1987).
2. S. John, Phys. Rev. Lett. **58**, 2486 (1987).
3. J. J. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals: Molding the Flow of Light* (Princeton University Press, Princeton 1995).
4. Y. Xia, B. Gates, Z.-Y. Li, Adv. Mater. **13**, 409 (2001).
5. S. Noda, K. Tomoda, N. Yamamoto, and A. Chutinan, Science **289**, 604 (2000).
6. V. Berger, O. Gauthier-Lafaye, and E. Costard, J. Appl. Phys. **82**, 60 (1997).
7. M. Campbell, D. N. Sharp, M. T. Harrison, R. G. Denning, and A. J. Turberfield, Nature **404**, 53 (2000).
8. S. Shoji and S. Kawata, Appl. Phys. Lett. **76**, 2668 (2000).
9. Yu. V. Miklyaev, D. C. Meisel, A. Blanco, G. von Freymann, K. Busch, W. Koch, C. Enkrich, M. Deubel, and M. Wegener, Appl. Phys. Lett. **82**, 1284 (2003).
10. D. C. Meisel, M. Wegener, and K. Busch, submitted.
11. E. Yablonovitch, T. J. Gmitter, and K. M. Leung, Phys. Rev. Lett. **67**, 2295 (1991).
12. S. Maruo, O. Nakamura, and S. Kawata, Opt. Lett. **22**, 132 (1997).
13. B. H. Cumpston, S. P. Ananthavel, S. Barlow, D. L. Dyer, J. E. Ehrlich, L. L. Erskine, A. A. Heikal, S. M. Kuebler, I.-Y. S. Lee, D. McCord-Maughon, J. Qin, H. Röckel, M. Rumi, X.-L. Wu, S. R. Marder, and J. W. Perry, Nature **398**, 51 (1999).
14. S. Kawata, H.-B. Sun, T. Tanaka, and K. Takada, Nature **412**, 697 (2001).
15. G. Witzgall, R. Vrijen, E. Yablonovitch, V. Doan, and B. J. Schwartz, Opt. Lett. **23**, 1745 (1998).

16. K. M. Ho, C. T. Chan, C. M. Soukoulis, R. Biswas, and M. Sigalas, Solid State Commun. **89**, 413 (1994).
17. H. S. Sözüer and J. P. Dowling, J. Mod. Optic. **41**, 231 (1994).
18. M. Deubel, G. von Freymann, M. Wegener, S. Pereira, K. Busch, and C. M. Soukoulis, Nat. Mater. in press.
19. O. Toader, M. Berciu, and S. John, Phys. Rev. Lett. **90**, 233901 (2003).
20. A. Blanco, E. Chomski, S. Grabtchak, M. Ibisate, S. John, S. W. Leonard, C. Lopez, F. Meseguer, H. Miguez, J. P. Mondia, G. A. Ozin, O. Toader, and H. M. van Driel, Nature **405**, 437 (2000).

# Photonic Crystal Waveguides: Dispersion, Anomalous Refraction and Applications

Remigius Zengerle

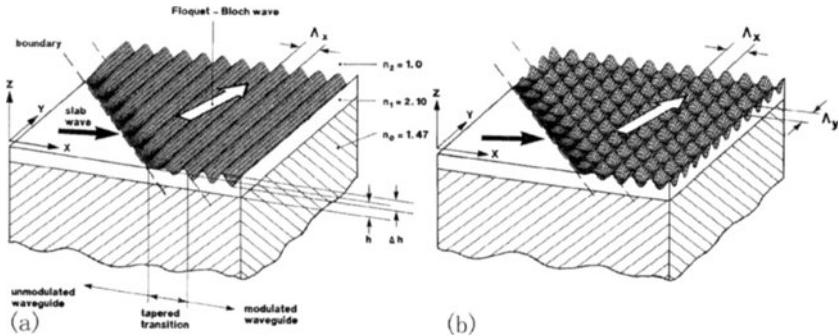
Technical University of Kaiserslautern  
Erwin-Schroedinger Street, 67663 Kaiserslautern  
[zengerle@rhrk.uni-kl.de](mailto:zengerle@rhrk.uni-kl.de)

**Abstract.** Early own experimental investigations on one- and two-dimensional strongly corrugated optical waveguides revealed both the existence of photonic bandgaps and anomalous refraction phenomena. Based on our experimental results, the dispersive and anomalous refraction properties of one- and two-dimensional photonic crystal waveguides are explained both phenomenologically and by simulations using Maxwells Theory. Possible new ways for applications of semiconductor photonic bandgap devices are also discussed.

## 1 Introduction

Periodic arrangements of – generally spoken – inhomogeneities in the distribution of matter have been investigated widespread in physics and used in engineering devices. Since the introduction of a special strongly modulated form, the so-called photonic crystals (PhCs) [1,2], these periodic structures have highly focused the activities of both the academic and the industrial communities. One special benefit of these structures is their ability to manipulate light at a scale of only several wavelengths. Although the initially proposed structure consisted of a three-dimensional periodic lattice [1], more recently two-dimensional PhC devices realized in planar slab waveguides have become a preferred area of research. One very common configuration of two-dimensional PhCs are semiconductor devices using different forms of integrated air-hole lattices. Many useful devices have been proposed [2] based on the additional introduction of irregularities in the periodicity, generally referred as "defects", allowing e.g. the realization of cavities and optical waveguides.

Photonic crystal structures can be used in two distinct ways: Direct use of the photonic bandgap itself for example as filters or for suppression of unwanted spontaneous emission. The other way is to use the strong dispersion [3] in the vicinity of an optical bandgap both in space or frequency domains. Especially the strong angular difference between phase and group (= energy) velocity may be used in practical devices. In surface acoustics, this difference is well known and denoted as the beam-steering angle [4].



**Fig. 1.** One- and two-dimensional periodic waveguides

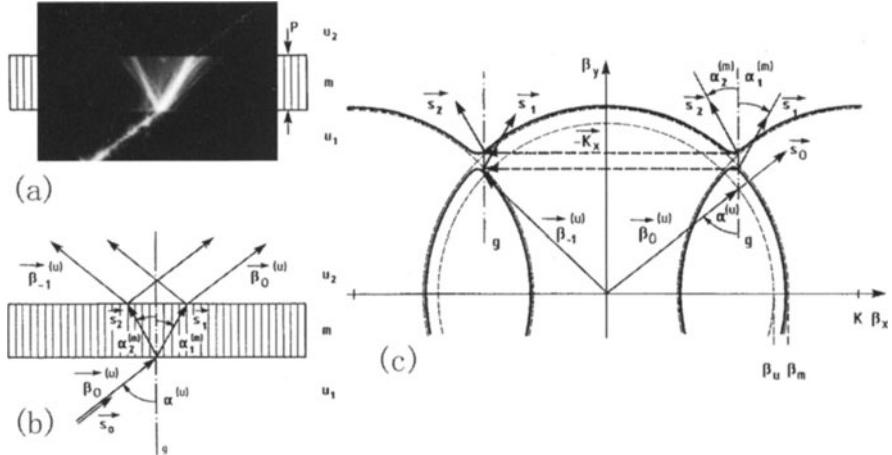
## 2 One- and Two- Dimensional Periodic Waveguides

Since the advent of integrated optics periodic waveguides played a central role in this technology. Investigations on the dispersive properties of metal meshes in infrared physics [5] induced a closer view at more-dimensional periodic light guides. So, wave propagation in singly and doubly periodic waveguides especially in the regime of Bragg Reflection was investigated in detail [6] and extended on periodic waveguides located in-between smooth waveguides, separated by straight transitions. So an intense study of diffraction and refraction phenomena at dielectric interfaces was possible, too [6,7,8]. A relatively high periodic modulation contrast was achieved using  $Ta_2O_5$  waveguides deposited on glass substrates. We investigated both singly and doubly periodic waveguides as shown in Fig. 1. Excellent grating quality was ensured by interferometric definition of the gratings.

## 3 Anomalous Refraction Phenomena in One-Dimensional Periodic Waveguides

### 3.1 Simultaneous Positive and Negative Ray Refraction

Anomalous refraction phenomena (Fig. 2) were demonstrated in a periodic waveguide with the effective index  $N^{(m)}$  in the modulated region being higher than  $N^{(u)}$  in the unmodulated region. The grating lines are oriented orthogonally to the boundaries. Wave propagation in the periodic waveguide is described by Floquet-Bloch waves consisting of an infinite set of spatial harmonics (SHs) interrelated by Floquet's theorem [9]. Due to the boundary conditions, two Floquet-Bloch waves corresponding to the upper and lower branches of the dispersion contour in the wavevector diagram (WVD) of the periodic structure [6] are excited simultaneously in the vicinity of Bragg-reflection. There they form two rays (indicating the trace of power flow, which is directed normally to the dispersion branches in the WVD) propagating in different directions (Fig. 2). This ray refraction can be described



**Fig. 2.** Simultaneous positive and negative ray refraction in a singly periodic planar waveguide. Grating orientation normal to the boundaries and  $\beta_m > \beta_n$  ( $\Lambda = 282.2$  nm,  $\lambda_0 = 666$  nm,  $p = 2$  mm). (a) Photograph of the propagation effects with enhanced contrast in the  $u_2$ -region. (b) Schematic ray representation c) WVD of the structure (TM polarization only, not according to scale). Arrows in the  $u$ -regions indicate the directions of the corresponding wave-vectors. 'g' denotes the normal to the waveguide boundaries

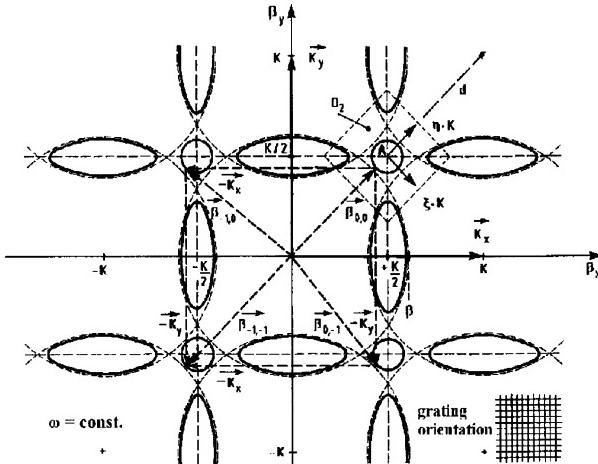
by two effective ray indices: one for "ordinary" refraction  $s_0$  into  $s_1$  with  $N_{r,1}^{(m)} > N^{(u)}$  and the other for the refraction  $s_0$  into  $s_2$  with a "negative refraction angle", thus leading to  $N_{r,2}^{(m)} < 0$ . In the upper transition to region  $u_2$  the Floquet-Bloch waves are reconverted into two sets of slab waves in the  $u_2$ -region, with the propagation constants  $\beta_0^{(u)}$  and  $\beta_{-1}^{(u)}$  according to the WVD in Fig. 2c.

#### 4 Refraction in Two-Dimensional Planar Photonic-Crystal Waveguides

Wave propagation in 2D periodic waveguides (Fig. 1b) is also described by Floquet-Bloch waves, consisting now in a doubly infinite set of spatial harmonics (SHs) interrelated by the two-dimensional extension of Floquet's theorem for a square grating [6,9]:

$$\beta_{\mu,\nu} = \beta_{0,0} + \mu K_x + \nu K_y \quad (1)$$

$K_x$  and  $K_y$  are the grating vectors of a square grating with  $K_x = K_y = K$ . Assuming wave propagation of one polarization only, the WVD now consists – at almost vanishing modulation – of a doubly infinite set of circles (each circle assigned to a special SH) as drawn in Fig. 3. The "Bragg frequency"  $\omega_B$  is reached if four neighboring circles intersect in one point, for example,



**Fig. 3.** Wave-vector diagram of a doubly periodic planar waveguide. Schematic diagram for one polarization only: - - - approximation circles; d: grating diagonal,  $\beta$ : propagation constant of the unmodulated waveguide

at  $\beta_x = \beta_y = K/2$ . In Fig. 3 the WVD is drawn for the normalized optical frequency  $\Omega = (\omega/\omega_B - 1) < 0$ , i.e. below the Bragg frequency. The dispersion branches split at the points of intersection of the circles, forming dispersion contours with strongly frequency-dependent shapes. In the actual case, an isolated nearly circular inner branch with frequency-dependent size exists which, however, even can disappear completely.

Doubly periodic planar structures, guiding one fundamental mode with TE and TM polarization, each, reveal rather complex dispersion characteristics near the regime of Bragg reflection. By selective excitation of Floquet-Bloch waves, such contours were measured in [8].

In order to gain easily insight into significant propagation phenomena, only TM-TM wave coupling is investigated here. The strongly frequency dependent shape of the dispersion contours in the WVD in the regime of Bragg reflection were calculated under simplifying assumptions (no z-dependence of the field components) using a convenient set of SHs.

Rigorous calculations using 16 SHs revealed that for the periodic modulation strengths of our experimental structures, only the amplitudes of four SHs (those with  $\beta_{0,0}$ ;  $\beta_{-1,0}$ ;  $\beta_{0,-1}$ ;  $\beta_{-1,-1}$ , see Fig. 3), may be significant. The amplitudes of all other SHs remain well below 1% of the strongest SH.

#### 4.1 Frequency Dependent Propagation in 2D Photonic Crystal Waveguides

Figure 4 illustrates at different optical frequencies various propagation phenomena observed in doubly periodic waveguides, embedded in-between two smooth waveguides with straight boundaries. The left column shows detail

D2 (view rotated 45°) of the WVD in Fig. 3 calculated according to an approximate contour equation [8]. The approximation circles are inserted as straight broken lines, the branch for the fundamental SH being emphasized. In the center row, a schematic ray-optical representation of wave propagation is given for some selected rays, whereas the right column photographs show the real propagation phenomena.

For studying the angular dependence of wave propagation at different frequencies, a divergent light beam is coupled by means of a prism into the unmodulated ( $u_1$ )-region of our doubly periodic test structure. The divergent guided beam (angular divergence about 50 mrad) is incident normally to the first boundary which is parallel to a diagonal of the grating.

## 4.2 Anomalous Super Refraction

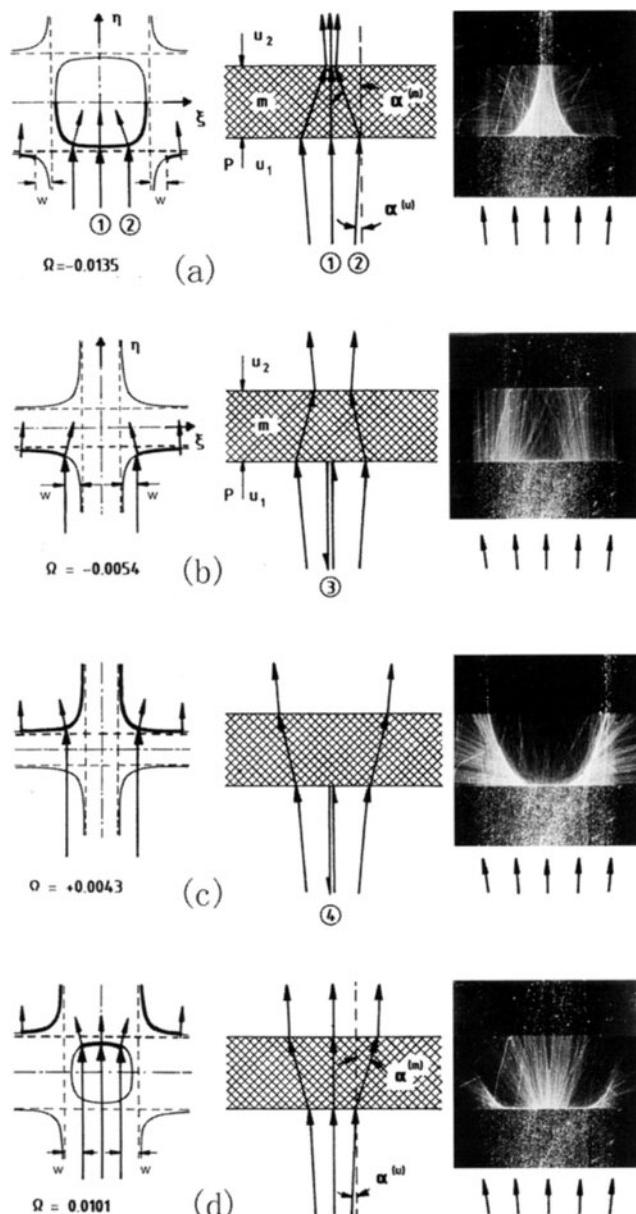
Figure 4a starts with propagation phenomena at a frequency  $\Omega$  well below the Bragg condition ( $\Omega = 0$ ). For ease of explanation, we use a simplified ray-optical picture where the incident divergent beam is separated into a series of rays with slightly divergent directions. The central normally incident ray "1" will excite, in the tapered transition, a Floquet-Bloch wave with a group velocity direction collinear to the incident ray. This leads to a straight passage through the areas  $u_1$ , m and  $u_2$ . In the WVD of Fig. 4a the lower half of the inner dispersion branch corresponds to a group-velocity flow directed upwards. Ray "2", however, inclined a small angle  $\alpha^{(u)}$  with respect to the normal of incidence, propagates, after conversion into a Floquet-Bloch wave, at the high deflection (beam-steering) angle  $\alpha^{(m)}$  in the photonic crystal waveguide (m-region). After passing the second boundary, the original ray direction is resumed. Owing to the concave shape of the inner dispersion branch in the WVD a slightly divergent beam is redirected towards the grating diagonal when passing across the doubly periodic structure.

Rays incident at angles significantly larger than the sketched angle  $\alpha^{(u)}$  will excite Floquet-Bloch waves at large beam-steering angles leading to a caustic curve similar to that of a high-aperture spherical mirror. Measurements on reflections of the incident light revealed that even at an angular group velocity deviation of 45° the reflection coefficient remained below 0.03.

This observed beam-focusing is somewhat unexpected in physics because it is achieved in a structure with straight-line geometry only (straight grating lines and straight boundaries). Further increase of the frequency will reduce the extension of the inner part of the dispersion contour even more which converges to the shape of a circle. The results are decreasing focal lengths even down to zero.

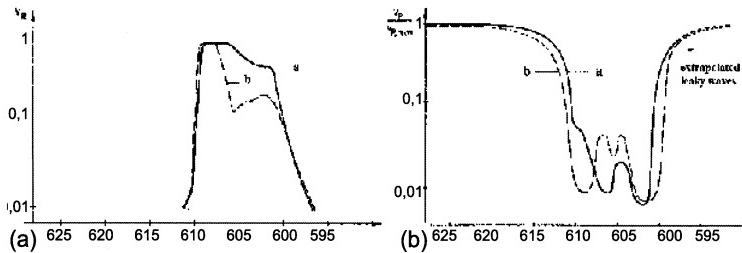
## 4.3 Photonic Bandgaps

The special case where the inner part of the dispersion contours disappears completely takes place for increasing frequencies beyond  $\Omega_l = -0.0054$



**Fig. 4.** Embedded two-dimensional photonic crystal waveguide. Dispersion phenomena at different optical frequencies using slightly divergent incident beams

(Fig. 4b). This absence of the inner dispersion branch holds in a frequency range up to  $\Omega = +0.0043$  (Fig. 4c). In this range rays incident within a small angular range around the normal of the boundary (which is identical to a diag-



**Fig. 5.** Measured photonic bandgap in a 2D photonic crystal waveguide. (a) Reflected power, (b) transmitted power

onal of the grating) are totally reflected, while outside this spectral stop-band no reflection should occur. Rays in the outermost angular regions of the incident beam, however, excite Floquet-Bloch waves in the outermost dispersion branches.

The existence of this photonic bandgap was confirmed by measurements [6]. A photonic bandgap requires a complete back-reflection of the incident waves into their original direction. This could be proven by introduction of a beam splitter. Figure 5 shows the normalized transmitted and reflected powers for two selected angles of incidence. The incident light beam with small angular divergence was either directed exactly into the grating diagonal (a) but also at a tilt of  $0.17^0$ (b).

#### 4.4 Normal Super Refraction

At frequencies  $\Omega > \Omega_1$ , the inner dispersion branch will appear again and Floquet-Bloch waves corresponding to the upper half of the dispersion characteristic (only these branches represent group-velocity components directed upwards) will be excited (Fig. 4d). The convex shape of the dispersion contour causes a strongly divergent beam in the periodic region, resulting in a beam expansion in region  $u_2$ . The amount of expansion is strongly frequency dependent and decreases with increasing optical frequency.

#### 4.5 Generalized Refraction Law

Beam steering of rays in doubly periodic waveguides with a straight boundary reveals special refraction phenomena. As shown in Fig. 4d, a ray incident at the angle  $\alpha^{(u)}$  is strongly refracted at an angle  $\alpha^{(m)}$  as it would be at the boundary of two media with highly different refractive indices, with the lower index being assigned to the grating region. On the other hand (Fig. 4a), the incident ray is also strongly refracted but in the opposite direction, establishing a strong 'negative' ray refraction. For both cases, a general refraction law for rays can be derived. According to [8 (16)] near  $\Omega = 0$  the relevant inner

part of the dispersion curves in the normalized WVD can be approximated by a circle of radius

$$\rho = [(\Omega^2 - \Omega_0^2)/2]^{1/2} \quad (2)$$

Neglecting the small additional refraction arising from differences of the average propagation constants in the unmodulated and the modulated regions, simple geometrical considerations lead to the following strongly frequency-dependent refraction law:

$$\sin \alpha^{(u)} / \sin \alpha^{(m)} = (\Omega^2 - \Omega_0^2)^{1/2} \operatorname{sgn}(\Omega) = N_r^{(m)} / N_r^{(u)} \quad (3)$$

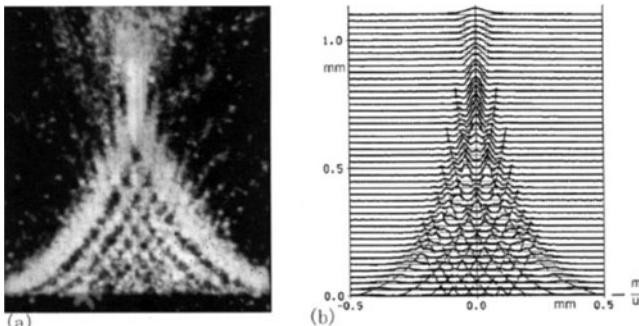
where 'negative' refraction is indicated by opposite signs of incidence and refraction angles. In analogy to Snell's law,  $N_r^{(u)}$  and  $N_r^{(m)}$  are effective ray indices of refraction.

#### 4.6 Interference of Floquet-Bloch Waves Excited by a Gaussian Beam

Detailed investigations on the intensity distribution of the scattered light in the cases of beam focusing at frequencies exceeding the case in Fig. 4a, reveal an inhomogeneous intensity distribution in the area between the caustic curves, as shown enlarged in Fig. 6. A pattern consisting of a series of bright curved stripes following the curvature of the outer two caustic curves with varying distance can be identified. These stripes may even intersect, forming a non-orthogonal two-dimensional intensity grating. This 'grating' can be identified as the interference pattern formed by the superposition of an infinite set of Floquet-Bloch waves excited by a divergent (Gaussian) beam incident at the boundary between smooth and corrugated regions.

In order to calculate numerically the interference pattern using given waveguide and beam parameters, we replaced the finite periodic planar waveguide by a periodic waveguide with infinite extension in the z-direction. The incident divergent beam with a Gaussian intensity distribution is decomposed in a quasi-infinite set of plane waves, propagating in slightly different directions. Passing the transition, these plane waves are converted into Floquet-Bloch waves. Each of them is represented with high accuracy by four SHs and individual amplitudes. All these plane SHs together form a very complex two-dimensional interference pattern with fringe spacings strongly varying with location.

This macroscopic fringe pattern can be calculated, using a convenient averaging procedure. Figure 6 reveals the good qualitative agreement between calculated and measured interference patterns, consisting of crossed interference fringes for a propagation path in the modulated region within the focal length, whereas no significant interference for greater path lengths takes place. Further investigations on Gaussian beams in periodic media are given in [10,11].



**Fig. 6.** Interference of focused Floquet-Bloch waves in a doubly periodic planar waveguide ( $\Omega = -0.0086$ , TM-TM wave coupling). A slightly divergent Gaussian light beam is incident onto the boundary (bl) between unmodulated ( $u$ -) and modulated ( $m$ -) regions. (a) Photograph of the spatial intensity distribution. (b) Calculated macroscopic intensity distribution. Characteristic maxima of the intensity patterns are interconnected by broken lines

#### 4.7 Negative Ray Refraction in State-of-the-Art III-V Semiconductor Waveguides

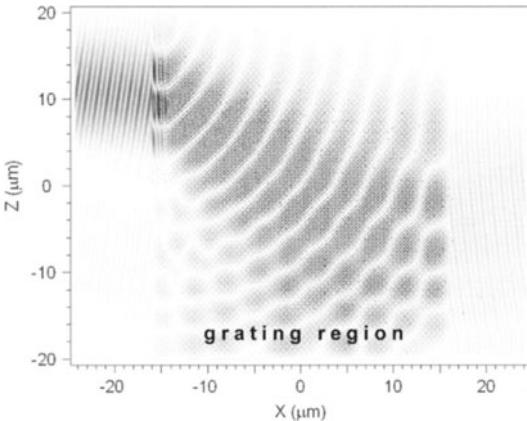
As example for present technological capabilities we use a 2D photonic crystal structure made of InGaAsP semiconductor cylinders, separated by air. The constants of the square grating were taken to be 282.2 nm as in our former experiments. The large refractive index of the quaternary material together with the diameter of the columns being in the range of half of the grating constant, leads to a very high effective index and thus the expected photonic bandgap increases into the range of 1100 nm to 1300 nm for propagation in the grating diagonal.

Figure 7 shows a Gaussian beam with a spot size of 10  $\mu\text{m}$  launched at  $z = -16 \mu\text{m}$  shortly in front of the periodic structure (located between  $-15 \mu\text{m} < x < +15 \mu\text{m}$ ) at an angle of  $87^\circ$  with respect to the vertical boundary. Only the resulting reflected and diffracted beams are shown.

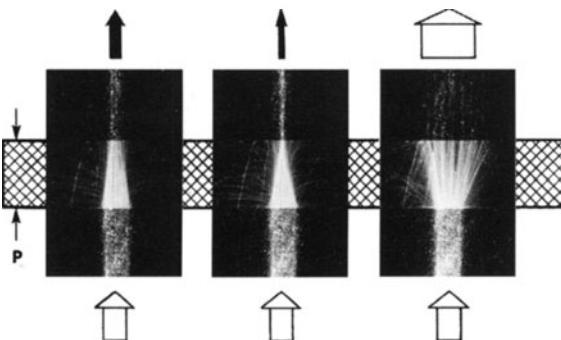
### 5 Applications

#### 5.1 Frequency-Dependent Spot-Size Transformation

Narrowing of the angular aperture of the incident beams in Fig. 4 by introducing an aperture near the first transition allows the demonstration of a strongly frequency-dependent spot-size transformer. The concentration ratio varies by about a factor of 3 within a relative frequency range of 0.16% only. The relative frequency difference between strong focusing and strong defocusing is in the range of 2%.



**Fig. 7.** Negative ray refraction shown for a Gaussian beam launched in front of a 2D photonic crystal waveguide. FDTD-simulation

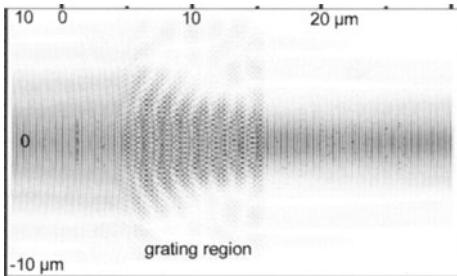


**Fig. 8.** Strongly frequency dependent spot-size transformation at  $\Omega = -0.0132$ ;  $-0.0116$ ;  $+0.0101$

## 5.2 Spot-Size Transformation in State-of-the-Art III-V Semiconductor Waveguides

Strongly modulated one- and two-dimensional semiconductor waveguide structures are used for novel functional devices [12]. One advantage of these structures is the ability of beam shaping even at a large angular divergence. Here we present the results of our newest theoretical investigations. For simplicity we took a square grating comprising dielectric columns with a refractive index of 3.2, surrounded by air, using our original grating constant of 282.2 nm. At a wavelength of 1158 nm we get a large nearly circular shape of the inner dispersion branch in the WVD in the direction of a grating diagonal below the lower cut-off frequency of the photonic bandgap.

The finite 2D-grating (9.9  $\mu\text{m}$  in length) was embedded in waveguides with an effective index of 2.3 and separated by straight boundaries. The square



**Fig. 9.** Spot-size transformation using a 2D photonic crystal waveguide

grating is oriented with one diagonal being orthogonal to the boundaries. A Gaussian beam (E-field being polarized orthogonally to the grating plane) was launched at  $z = -46 \mu\text{m}$  in front of the first boundary (at  $z = 4 \mu\text{m}$ ). The launching spot-size (beam waist) was taken to be  $3.7 \mu\text{m}$ . This leads to a spreading of the mode-field diameter up to  $9.5 \mu\text{m}$  in front of the first boundary. Due to strong negative refraction in the photonic crystal structure the propagating beam is refocused in the periodic region leaving it at the second boundary with a spot-size of  $3.6 \mu\text{m}$ .

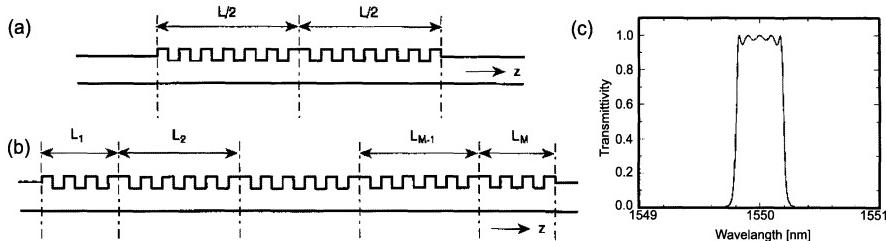
### 5.3 Further Applications

The strong birefringence in photonic crystal waveguides near the photonic bandgap enables applications like frequency filtering and polarization splitting. We used the corresponding strong polarization dependent beam steering in a 1D photonic crystal waveguide for efficient separation of transmission bands [13] and TE-TM polarization splitting [14] with a minimum extinction ratio of 17 dB.

Doping of photonic crystals is a well known method for tailoring its spectral behavior. For filter applications, the pass-band should show a so-called "flat-top" with steep slopes to the stop bands. By extending the design of the so-called high-Q transmission filter [15] we were able to verify numerically, that inhomogeneous multiple doping of a 1D photonic crystal waveguide using several phase-shift regions may lead to the desired filtering characteristics [16]. An optimized choice of individual subsection lengths (Fig. 10c) resulted in a residual pass-band ripple below 0.3 dB.

## 6 Conclusions

Photonic crystal waveguides permit applications of a large variety of effects, both in the photonic bandgap region itself and in its vicinity. The major application of photonic bandgaps is the suppression of unwanted frequency bands whereas strong dispersion phenomena including negative refraction



**Fig. 10.** Phase shifted Bragg-grating filters with improved transmission characteristics. (a) Singly-doped photonic crystal waveguide. (b) Multiply-doped 1D-photonic crystal filter. (c) Transmission spectrum for 5 phase-shift regions of different lengths

occur in the vicinity of photonic bandgaps. We were able to verify most of the suitable effects already in waveguides with only moderate one- and two-dimensional periodic modulation depth. Potential applications using state-of-the-art semiconductor devices were shown by FDTD-simulations.

### Acknowledgements

The author wants to thank Reinhard Ulrich for suggesting and supporting the fascinating work in the field of photonic bandgap structures already in a very early phase.

### References

1. E. Yablonovitch, Phys. Rev. Lett. **58**, 2059 (1987).
2. J. Joannopoulos et al, *Photonic Crystals*(Princeton Press, Princeton, N.J. 1995).
3. P. St. J. Russell, Appl. Phys. B **26**, 37 (1981).
4. I. M. Mason and E. A. Ash, J. Appl. Phys. **42**, 5343 (1970).
5. R. Ulrich, M. Tacke, Appl. Phys. Lett. **22**, 251 (1973).
6. R. Zengerle, PhD Thesis, University of Stuttgart (1979).
7. R. Ulrich, R. Zengerle, (Topical Meeting on Integrated and Guided-Wave Optics, Incline Village, paper TuB1 1980).
8. R. Zengerle, Journal of Modern Optics **34** 1589 (1987).
9. L. Brillouin, *Wave propagation in Periodic Structures* (McGraw-Hill, New York 1946).
10. R. Zengerle, O. Leminger, IOOC 89, Kobe, (1989).
11. R. Zengerle, O. Leminger, URSI International Symposium on Electromagnetic Theory, Sydney, 19 (1992).
12. T. D. Happ, M.Kamp, A. Forchel et al., Appl. Phys. Lett. **82**, 4 (2003).
13. R. Zengerle and O. Leminger, J. Opt. Commun. **11**, 11 (1990).
14. R. Zengerle, Electronics Lett. **24**, 11 (1988).
15. H. A. Haus: *Waves and Fields in Optoelectronics* (Prentice-Hall, Englewood Cliffs 1984).
16. R. Zengerle and O. Leminger, Journal of Lightwave Technology **13**, 2354 (1995).

# Photonic Crystal Structures in Ultrafast Optics

A. Tünnermann<sup>1,3</sup>, T. Schreiber<sup>1</sup>, M. Augustin<sup>1</sup>, J. Limpert<sup>1</sup>, M. Will<sup>1</sup>,  
S. Nolte<sup>1</sup>, H. Zellmer<sup>1</sup>, R. Iliew<sup>2</sup>, U. Peschel<sup>2</sup>, and F. Lederer<sup>2</sup>

<sup>1</sup> Friedrich-Schiller-Universität Jena

Institut für Angewandte Physik, Max-Wien-Platz 1, 07743 Jena

<sup>2</sup> Friedrich-Schiller-Universität Jena

Institut für Theoretische Optik, Max-Wien-Platz 1, 07743 Jena

<sup>3</sup> Fraunhofer Institut für angewandte Optik und Feinmechanik

Albert-Einstein-Str. 7, 07745 Jena

**Abstract.** During the past several years the study of light propagation in photonic crystals has attracted a steadily growing interest. Here we evaluate the potential of low index contrast photonic crystals realized in film waveguides and fibers. Based on the characterization of these structures (attenuation, mode profiles and dispersion) we discuss various applications in ultrafast optics.

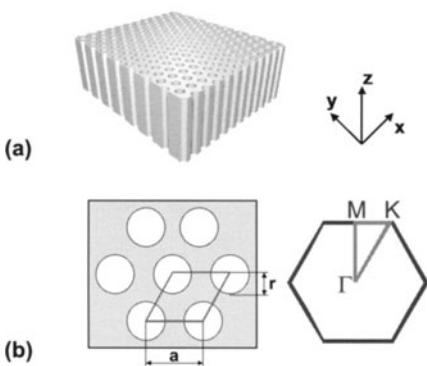
## 1 Introduction

Within the last decade light propagation in photonic crystals has been extensively investigated. This research was primarily driven by the foreseeable potential of photonic crystals in miniaturized optoelectronic devices required in many fields and in particular in optical communications [1]. It is now widely appreciated that photonic crystals represent an artificial optical material where the flow of photons can be controlled in many aspects, similar to the flow of electrons in bulk or low dimensional semiconductors [2]. On the one hand photonic crystals can exhibit band gaps, like semiconductors do, where basically light propagation in a certain wavelength domain in all or specific directions is prohibited. On the other hand light that may propagate in these structures or at least along intentionally introduced defects usually exhibits particular features, e.g., field profiles, dispersion and diffraction properties can be tailored to a large extent. A further inspiration to fabricate these materials came from their existence in nature, for instance as opals, at the wings of some butterflies and in the hairs of the sea mouse. For some insects they even provide biological functions [3]. Nowadays photonic crystals are fabricated on the basis of various dielectric and semiconductor materials where the refractive index of the host material is an important parameter. By changing the structure of the photonic crystal the properties can be controlled. A simple example of an one-dimensional photonic crystal is the dielectric multilayer film, e.g. a Bragg stack, whereas two-dimensional photonic crystals exhibit a periodicity in a plane perpendicular to the structure. Practical

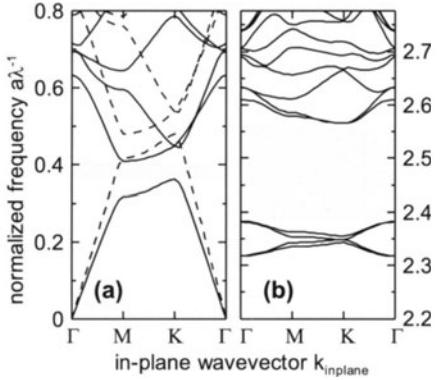
realizations are photonic crystal fibers [4] and structures with a huge aspect ratio, e.g. two-dimensional photonic crystal slabs. Even three-dimensional photonic crystals, which are harder to realize, have been fabricated and are a subject of contemporary research [5].

The engineerable optical properties of photonic crystals permit the control of light by tailoring the respective structures. Various novel properties can be exploited in ultrafast optics. Typical devices to be used are high quality and miniaturized cavities, switches, add-drop filters and efficient waveguide bends and splitters. By exploiting the peculiar dispersive properties of photonic crystals several devices such as optical delay lines (slow light), dispersion compensators for short pulse generation and white light generators have been developed. Spatial effects can be used for wavelength-selective functions, such as WDM filters and multiplexers allowing extraordinary angle and wavelength-sensitive light propagation through a photonic crystal (super prism). Even anomalous refraction and diffraction has been proved in photonic crystals.

This paper focuses on the two-dimensional case, namely photonic crystal slab waveguides (PCWG) and photonic crystal fibers (PCF). The theoretical description of photonic crystals is generally based on modified methods of electronic band structure theory [1,6,7]. Its understanding is essential for characterizing existing as well as designing novel structures. Fig. 1 shows a schematic of a two-dimensional photonic lattice. The necessary refractive index contrast is achieved by hexagonally arranged air-filled holes in a transparent material; different periodic configurations are also possible. The unit cell of this lattice is characterized by the size  $a$  and the radius of the holes  $r$ . The propagation directions in the plane of periodicity are referred to as the wave vectors of the corresponding Brillouin zone. The propagation properties are summarized in a band gap structure shown in Fig. 2a for in-plane propagation ( $xy$ -plane) (Details of the parameters are given in the subtitle of the figure of simplicity.) For a high index contrast ( $\epsilon=3.8$ ) a full photonic band gap exists for TE polarization, where the light is not allowed to propagate in-plane. The physical reason is, that at these wavelengths light is multiply scattered and interferes destructively. Introducing defects into this perfect periodic crystal leads to localized modes in much the same way as substituted atoms define the properties of semiconductor bandgaps. In photonic crystals this can be done by removing several columns from the lattice or by changing their hole diameter. This perturbation breaks the symmetry and discrete modes located inside the band gap are possible, which means that light is allowed to propagate inside the defect but cannot penetrate into the surrounding crystal. Photonic crystal waveguides are based on this idea. Removing a line from the lattice and additionally confining the light in the  $xy$ -plane by total internal reflection in a waveguide film forms the waveguide. The fabrication, properties and potential application of highly efficient PCWG are described in detail in section 2



**Fig. 1.** (a) Two-dimensional photonic crystal formed by a periodic array of holes in a transparent material. (b) The parameters are  $a$  as the size of the unit cell,  $r$  the radius of the holes and  $\Gamma$ , M and K as the high-symmetry points of the corresponding Brillouin zone

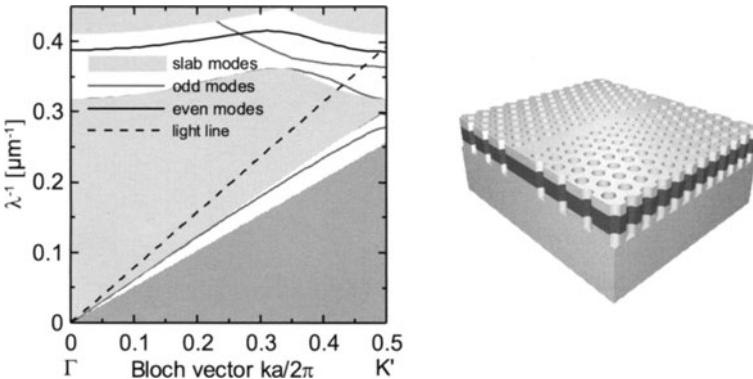


**Fig. 2.** (a) Photonic TE bandgap structure for  $r/a=0.3$ ,  $\beta_z=0$  and  $\epsilon=3.8$  showing a full bandgap (solid line). The bandgap disappears for  $\epsilon=2.1$  (dashed line). (b) bandgap structure corresponding to  $\epsilon=2.1$ ,  $\beta_z/G=2.5$  where  $G=2\pi/a$

So far, PCWG deal with in-plane propagation. If the index contrast of the perfect photonic crystal is decreased, e.g. by using fused silica ( $\epsilon=2.1$ ), the band gap for in plane propagation will disappear (Fig. 2a, dashed line). But for propagation perpendicular to the periodic plane a bandgap still exists as is shown in the band structure in Fig. 2b. Again, a defect can lead to a localized mode inside this bandgap and allows the propagation along the defect. This is the idea of photonic bandgap fibers (PBF), where light can be trapped and guided. If the defect is an increased air-filled hole, light cannot escape into the cladding and can be guided in air without diffraction. With such a fiber new applications are possible and some limits in modern optics can be overcome. For instance, air-core bandgap fibers have the potential of a lower loss than conventional fibers, where Rayleigh scattering limits the minimum loss in fused silica. Also, light interaction with gas molecules usually limited by the Rayleigh length of a focused beam is increased when the light is guided in a PBF. In ultrafast optics the low interaction of light with the host material and the unusual temporal dispersion properties can be used. Some of such applications of air-core bandgap fibers will be described by two examples in section 3

## 2 Two-Dimensional Photonic Crystal Slab Waveguides

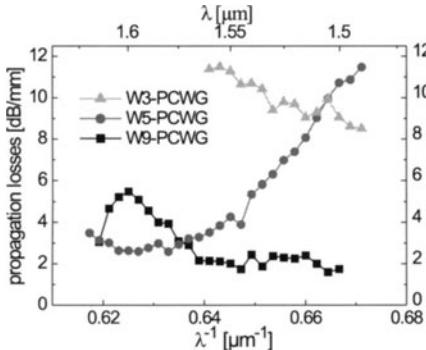
As mentioned above two-dimensional photonic crystals can be realized in slab systems where a high index waveguiding film is sandwiched between low index substrate and cladding in order to confine the light in the lattice plane



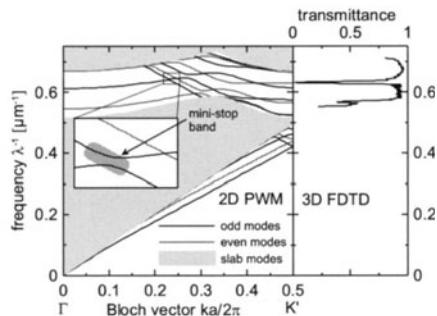
**Fig. 3.** Bandgap structure for a one line defect (W1, 2D slab from Fig. 1b with  $r/a=0.3$ ,  $\varepsilon=3.8$ , dark grey: evanescent modes in core

( $xy$ ) by total internal reflection. The propagation properties inside this plane are tailored by periodic structuring. Waveguides are realized by omitting one or more rows of holes. In our experiments we used a slab system consisting of a waveguiding film of 500 nm of  $\text{Nb}_2\text{O}_5$  ( $n=2.2$ ), a 300 nm cladding layer and a 2000 nm buffer layer both made of  $\text{SiO}_2$  ( $n=1.43$ ). The  $\text{SiO}_2$  buffer layer is located on a silicon substrate in order to produce optical facets by cleaving the sample.  $\text{Nb}_2\text{O}_5$  as the waveguiding film was chosen because its transparent region extends into the visible allowing the realization of photonic crystals in this spectral range.

The photonic crystal waveguides discussed here were designed by 2D and 3D band structure calculations, using a freely available software package [7]. A complete band gap for TE polarization is obtained for a hexagonal arrangement of holes with a diameter of 390 nm at a pitch of 620 nm ( $r/a=0.31$ ). By omitting one row of holes along the direction of the crystal so-called W1 defect channels inside these PCs were formed. Figure 3 shows the band structure for such a W1 PCWG. The white regions mark wavelengths where the propagation into the photonic crystal is forbidden. Propagation into the photonic crystal is only possible for wavelengths lying in the gray regions. As can be seen in Fig. 3 between  $0.35$  and  $0.41 \mu\text{m}^{-1}$ , a complete band gap for TE polarization is obtained. The waveguiding modes are shown by the red and black lines indicating the odd and even modes, respectively. Even and odd refer hereby to the parity with respect to the symmetry plane perpendicular to the layers and parallel to the waveguide. It is shown in Fig. 3 that waveguiding along the defect channel can occur for wavelengths lying inside the photonic band gap. The dashed line in Fig. 3 indicates the so-called light line. Since the effective index for modes located above the light line is smaller than the index of the substrate the light is no longer confined in the plane by total internal reflection and out-of-plane scattering losses occur. In this case the intrinsic losses (not taking into account the material absorption and



**Fig. 4.** Wavelength dependence of the propagation losses shown for different PCWGs



**Fig. 5.** 2D plane wave calculation (left) and 3D-FDTD obtained transmission spectra (right) of the W5-PCWG (hole diameter 380 nm, pitch 620 nm)

material scattering) of W1-PCWGs are typically in the range of 100 dB/mm to 300 dB/mm [8]. However, the propagation losses of wider photonic crystal waveguides (W3, W5, ...) are expected to be significantly smaller. Therefore, we realized W3, W5 and W9 waveguides with a length of 10 mm using e-beam lithography [13,14]. In contrast to W1-PCWGs these wider PCWGs are all multi-mode. In order to determine the propagation losses in these devices the transmission is measured for different sample lengths (so-called cut-back method). A linear fit yields both the propagation and the coupling losses. In Fig. 4 the wavelength dependence of the propagation losses of the W3, W5 and W9 are summarized. The lowest propagation losses for the PCWGs are measured to 8.5 dB/mm, 2.6 dB/mm and 1.7 dB/mm, respectively. These propagation losses are considerably low even when compared with results published for semiconductor PC structures [9,10,11,12].

Further characterization was done by determining the near-field intensity distribution at the output facets of the W3, W5 and W9-PCWGs. After a propagation distance of 60  $\mu\text{m}$  all PCWGs show a multi-mode profile in accordance with the band structure calculations. However, after a propagation distance of 0.8 mm for the W3 and W5 only one mode can be observed, while the W9 still shows two modes. After a propagation distance of 2.3 mm even the W9 shows only one remaining mode, i.e. after a sufficiently large propagation distance all PCWGs are effectively single mode. This is caused by the deeper penetration of higher order modes into the photonic crystal, which thus experience higher out-of-plane scattering losses [13,14]. Due to the effective single-mode guidance and the low transmission losses these wide waveguides provide an interesting basis for the realization of various devices. As examples, filtering devices and efficient waveguide bends will be demonstrated in the following.

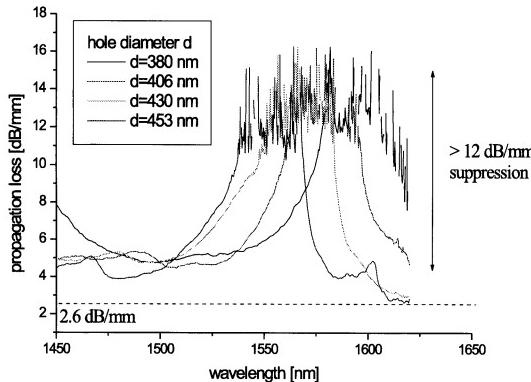
## 2.1 Filtering Devices Based on Photonic Crystal Waveguides

A filtering device based on a defect channel waveguide can be obtained by exploiting the occurrence of so-called mini-stop bands. Mini-stop bands (MSB) are mode-coupling phenomena, where the energy of the fundamental mode is transferred to higher order modes, resulting in much higher propagation losses and therefore producing a dip in the transmission spectrum. This property can be used as a filter to block wavelengths lying inside the bandwidth of the MSB. For a W5-PCWG consisting of air holes with a diameter of 390 nm and a lattice pitch of 620 nm the location of the MSB has been determined at 1560 nm ( $\lambda^{-1}=0.641 \mu\text{m}^{-1}$ , see Fig. 5 left). Several photonic crystal structures with a lattice pitch of 620 nm and hole diameters ranging from 380 to 453 nm have been fabricated with electron beam lithography. Scanning electron microscope (SEM) images of the optical facets show deeply etched holes with a depth of 1200 nm.

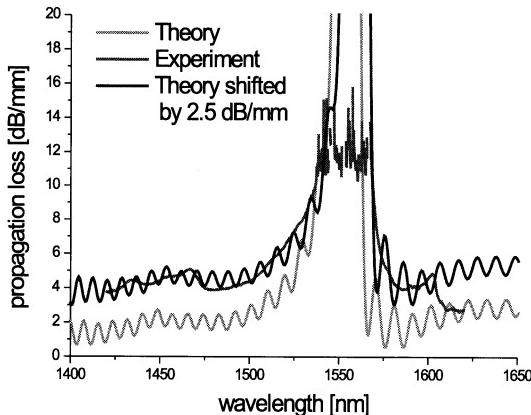
After cleaving the sample, W5-PCWGs with a length of 4 mm were obtained. The polarized light of a super-luminescence diode emitting between 1420 and 1620 nm was coupled into the waveguides via a long-distance microscope objective (125x) with a high numerical aperture of 0.8. At the output facet the light was collected by another microscope objective (40x, NA=0.65) and coupled into a fiber connected to an optical spectrum analyzer. The experiment was repeated without the sample in order to normalize the transmission spectra of the PCWGs. To obtain the propagation loss of the PCWGs, the stray light was detected for one fixed wavelength with a microscope objective (125x, NA=0.8) and an infrared camera at different positions along the waveguide. By fitting the data with an exponential decay function the propagation loss could be obtained. This value was used to normalize the propagation loss for the whole transmission spectra.

Figure 6 shows the measured wavelength dependence of the propagation loss for W5-PCWGs with different hole diameters. A shift of the MSB (corresponding to the peak of highest propagation loss) to higher wavelength with decreasing hole diameter is clearly observable.

The lowest propagation loss could be determined to 2.6 dB/mm at a wavelength of 1620 nm for the W5-PCWG with a hole diameter of 453 nm. The absolute MSB suppression was too high to be measurable; after 4 mm propagation only background noise was observed. In order to compare the accuracy of the 3D-FDTD and to be able to estimate the suppression of the MSB, 3D-FDTD simulations were performed for the W5-PCWG with 620 nm lattice pitch and 453 nm hole diameter. Figure 7 shows an excellent agreement in MSB location between experiment and theory. However, the theory underestimates the measured losses by approx. 2.5 dB/mm. Since the slab losses amount to only  $\sim 0.4$  dB/mm the majority of these losses is probably due to residues of the polymer-etching mask resulting in higher scattering losses. Based on these 3D-FDTD calculations a MSB suppression of approx. 200 dB/mm can be expected. Taking in account TE propagation losses of



**Fig. 6.** Wavelength dependence of the propagation loss for W5 PCWG with different hole diameters ranging from 380 nm to 453 nm. A shift of the position of the mini-stop band in dependence of the hole diameter could be detected

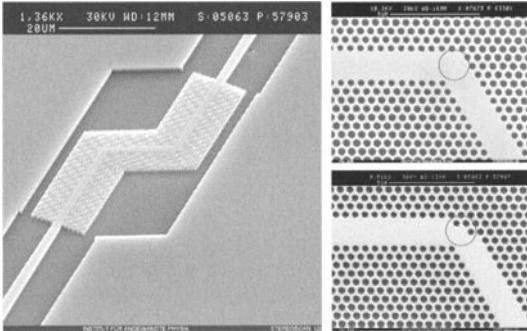


**Fig. 7.** Comparison between calculated and measured propagation loss of W5 PCWG with a hole diameter of 453 nm. The fringes in the theoretical spectrum are due to reflections at the borders of the computational window

the W5 of approx. 3–4 dB/mm, even 10 times shorter waveguides ( $\sim 100 \mu\text{m}$ ) still produce a significant suppression of  $\sim 20$  dB while outside the MSB spectral range only moderate losses of  $\sim 0.4$  dB are obtained. At this length the FWHM-bandwidth of the MSB amounts to  $\sim 15$  nm. With this performance the application of W5-PCWGs as a filtering device with a narrow bandwidth and high suppression is possible.

## 2.2 Photonic Crystal Waveguide Bends

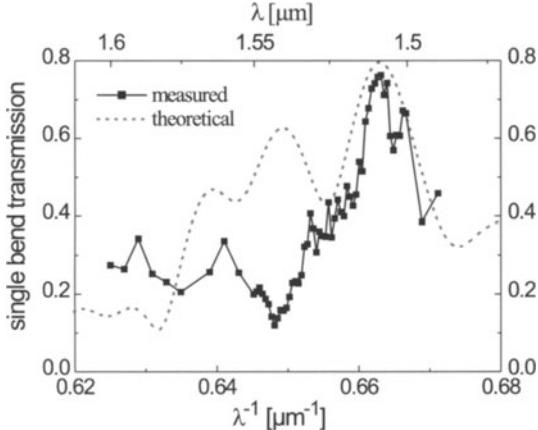
In addition to filtering devices another prominent example are photonic crystal waveguide bends which allow to redirect light on an extremely small length scale. Such bends are formed inside the PC by connecting two W3-



**Fig. 8.** SEM-images of the W3-PCWG double bend (left) with the optimized bend design (bottom right). The conventional bend design is also shown (top right). The PC consists of holes with a diameter of 374 nm and a depth of 1.1  $\mu$ m at a lattice pitch of 595 nm

PCWGs along the K directions. We realized a PC double bend by arranging two of these 60°-bends consecutively. The overall length of the PC structure was limited to 60  $\mu$ m in order to monitor the effects of the photonic crystal waveguide bend and to avoid suppression of higher order modes during propagation through the photonic crystal waveguide. By optimizing the bend design [14] (three extra holes are added at the bend section) a transmission of up to 80% can be achieved. Figure 8 shows a realized W3-PCWG double bend with a total length of 60  $\mu$ m. For transmission normalization purposes also straight PCWGs with the same length were fabricated. SEM analysis of the produced devices shows holes of 1100 nm depth and with an average diameter of 374 nm at a separation of 595 nm. The wall angles amount to 85°, and an aspect ratio of 1:3 is achieved. Furthermore, for easier handling of these PCWGs ridge waveguide tapers were used for coupling into and out of the photonic crystal bends.

For obtaining the bend efficiency the transmission through the photonic crystal double bend was measured. The bend transmission (efficiency) of a single bend was obtained by calculating the square root of the double bend transmission normalized on the transmission of a straight PCWG. We determined a maximum transmission of 75%/bend at 1515 nm. In Fig. 9 the experimental values for a single bend are compared to theoretical predictions based on 3D-FDTD simulations. Although some deviations between the experimental data and the FDTD-simulations exist, both the maximum transmission per bend and the position of this peak are in very good agreement with the theoretical predictions. These values promise the use of photonic crystal structures with low index contrast for various applications.

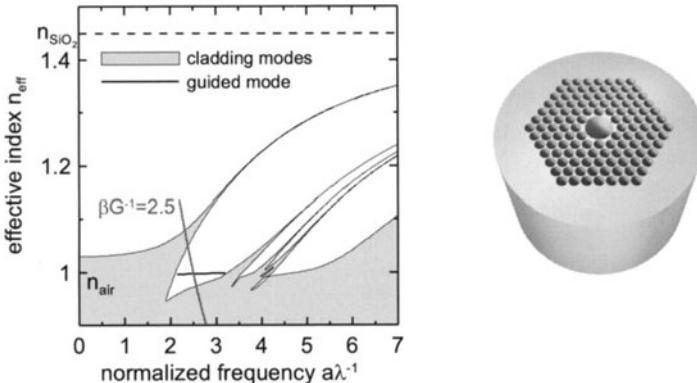


**Fig. 9.** Measured bend efficiencies for the optimized bend in comparison with 3D-FDTD simulations

### 3 Air-Core Photonic Band Gap Fibers

As mentioned in the introduction, the band gap disappears in a hexagonal photonic crystal for a low index contrast (Fig. 2a, dashed line). In propagation perpendicular to the periodic plane a bandgap still exists. This is the case for a fiber, where the holes of the crystal structure run down along the fiber length. Figure 10 shows the reduced bandgap structure for all possible propagation constants (effective index) in  $z$ -direction. The full bandgap diagram from Fig. 2b is indicated as a red line, corresponding to a wave vector of  $\beta_z/G=2.5$  where  $G=2\pi/a$  is the size of the first Brillouin zone. If a defect is introduced, e.g. a hollow core with the size of twice the hole diameter, a mode can exist inside the band gap and light is guided in air (Fig. 10, black line).

Guiding light in such an air-core requires the fabrication of photonic crystal fibers with large air-filling fraction to achieve a photonic band gap with the effective index of the defect mode below 1.0. Then light is guided in a well-defined wavelength range and trapped by a 2D photonic band gap of the cladding [15,16,17]. Due to the confinement of the radiation in the hollow core the nonlinearity is significantly reduced. Experiments have shown that the nonlinear coefficient of the hollow core fiber mode is very close to that of air [18]. The dispersion characteristic of such a bandgap fiber with a well-defined transmission range is in first order determined by the Kramers-Kronig relation and the waveguide dispersion. Therefore, the dispersion is anomalous over much of the transmission band. Anomalous dispersion implies that such fibers can support optical solitons. This was recently demonstrated by the soliton-type propagation of 5 MW peak power pulses in the 1.5  $\mu\text{m}$  wavelength region in a 1.7 m long bandgap fiber [19]. This high peak power was possible due to the very low nonlinearity in the air-core. In a conventional

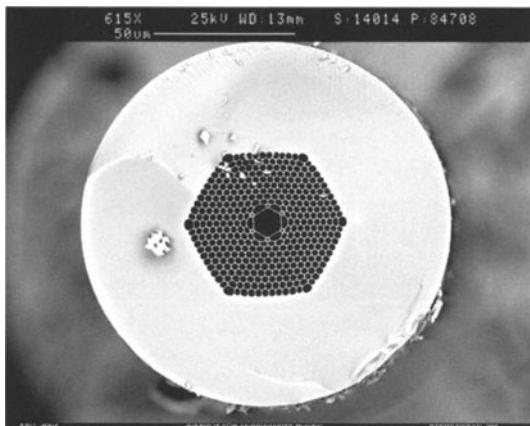


**Fig. 10.** Effective index diagram for propagation in direction along the defect with an air filling fraction of 92.5%, air filled defect radius  $2r$  and  $\varepsilon=2.1$ . The bandgap structure from Fig. 2 is indicated as a red solid line. A guided mode exists even in the air core (black line)

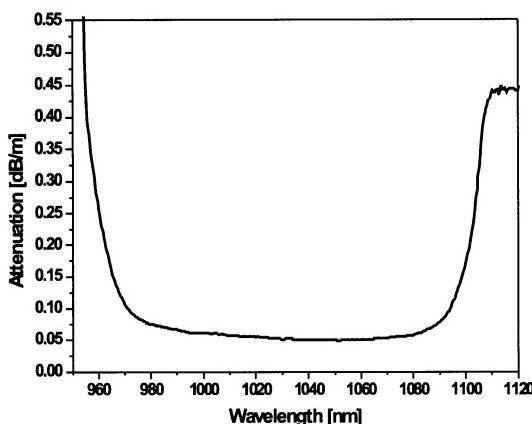
step-index single mode fiber the propagation of such pulses over comparable fiber length would be impossible due to enhanced nonlinear pulse distortion. Before the practical realization of real photonic bandgap fibers, holey fibers, which guide light by modified total internal reflection, already possess remarkable new features. Since recently the fabrication of fibers is possible which guide light by a photonic bandgap effect. Band gap fibers with guidance in a solid core [15] and even low loss hollow core fibers [20] were fabricated and different experiments already showed their potential for application [21,22]. Here we describe two experiments to show the potential of hollow core fibers in the field of ultrafast optics. After a short analysis of the main optical properties of the fiber, high peak-power fiber-based beam delivery and an all-fiber chirped-pulse-amplification (CPA) system based on compression in an air-core PCF is demonstrated.

### 3.1 Properties of an Air-Core PCF

The air-silica photonic bandgap fiber used in our experiments (Crystal Fibre AIR-10-1060) is shown in Fig. 11. It has a core diameter of 10.5  $\mu\text{m}$  and the bandgap ranges from 980 nm to 1080 nm with an attenuation of less than 0.1 dB/m. The spectral attenuation of this fiber is shown in Fig. 12. The group-velocity parameter at a wavelength of 1040 nm is experimentally determined to  $-0.054 \text{ ps}^2/\text{m}$  by measuring the temporal pulse broadening of an initially unchirped pulse in the fiber. The propagation is assumed to be linear (only dispersion), since no changes in the spectrum are observable. Theory suggests that at the corners of the bandgap the dispersion is increased dramatically. It also indicates that ultrashort pulse propagation may therefore be limited by higher order dispersion not measured here. Experimentally, the



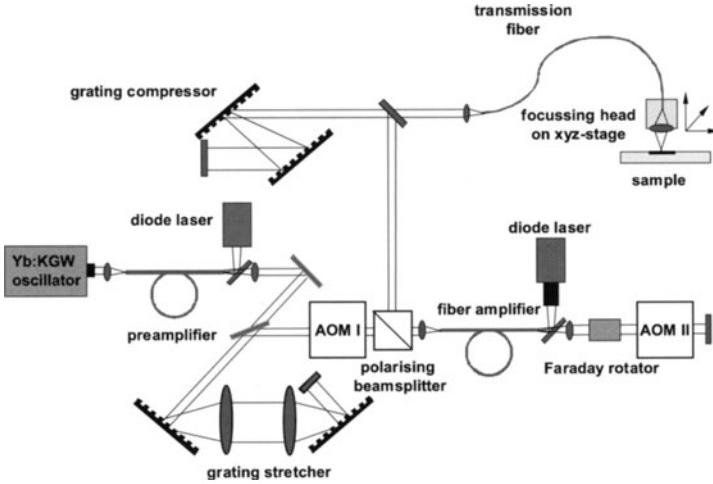
**Fig. 11.** Microscopic image of the air-core photonic bandgap fiber (Crystal Fibre AIR-10-1060)



**Fig. 12.** Measured spectral attenuation of the air-guiding photonic bandgap fiber

fiber supports only one fundamental mode with a Gaussian like mode profile. The numerical aperture is  $NA=0.12$ .

The low nonlinearity and the anomalous dispersion properties are very attractive for several applications in ultrafast optics as will be shown. Furthermore, the risk of damage is reduced due to the air guiding. Two examples will be given in the next paragraphs. Firstly, the fiber delivery of megawatt peak power femtosecond pulses and their application to micro-machining is demonstrated. Secondly, the anomalous dispersion properties and the low nonlinearity is used to replace a grating compressor in an all fiber chirped pulse amplification system.

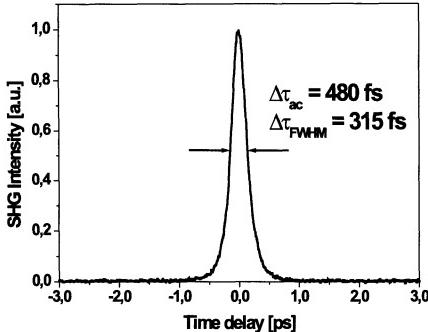


**Fig. 13.** Principle of the fiber CPA system and the air-core photonic bandgap fiber used for pulse transmission and micro-machining

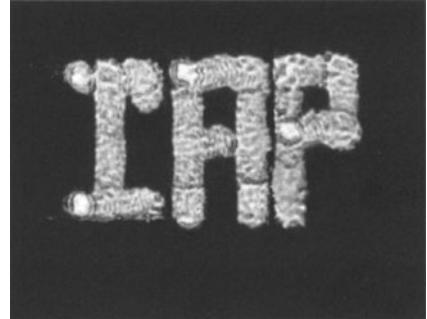
### 3.2 Megawatt Peak Power fs Pulse Delivery through Air-Core PCFs

Usually it is impractical to fiber deliver high-energy pulses due to extreme nonlinear pulse distortion even in large mode area fibers. This task is even more challenging for ultrashort pulses where the peak power can exceed several megawatt. For many applications such a flexible transport of radiation would be advantageous. However, air guiding photonic bandgap fibers have the potential for high power transmission [17].

The experimental setup of a fiber CPA system delivering high repetition rate  $\mu\text{J}$  femtosecond pulses is shown in Fig. 13. The seed source is a passively mode-locked femtosecond Yb:KGW oscillator. The fiber amplifiers are constructed using an air-clad microstructured ytterbium-doped large-mode-area fiber [23]. The repetition rate can be adjusted according to the application requirements using acousto-optical modulators (AOM) in a wide range from 10 kHz to several MHz. The recompressed pulses are coupled into the air-core PCF. Figure 14 shows the autocorrelation trace after the fiber. At an average power of 120 mW and a repetition rate of 115 kHz, which corresponds to  $1.04\mu\text{J}$  pulse energy the pulse duration out of the air-core fiber is 315 fs assuming a  $\text{sech}^2$ -pulse shape. The peak power directly out of this fiber is as high as 3.3 MW. This experiment demonstrates that the use of an air guiding photonic bandgap fiber allows the delivery of femtosecond pulses with  $\mu\text{J}$  energy. In a first application these pulses are used to microstructure a chrome layer on glass. The promising result is shown in Fig. 15. In the future this system will be used for the fabrication of photo-induced waveguides.



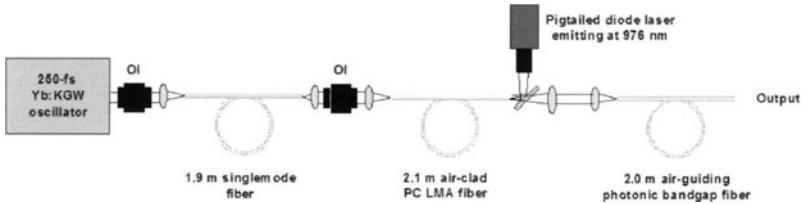
**Fig. 14.** Autocorrelation of the femtosecond pulse after 2 m fiber length



**Fig. 15.** Microscope image of a microstructured chrome layer on glass (back-light illumination)

### 3.3 All Fiber CPA System Using Compression in an Air-Core PCF

High energy and high repetition rate femtosecond pulses are required by different applications in medicine [24], micro- and nano- machining of solids [25], or the fabrication of photonic devices in dielectrics [26]. Applying the CPA technique [27] in fiber amplifiers is very attractive for constructing compact and practical high-energy [28] and high average power femtosecond pulse sources [29]. As shown in the previous section, in its simplest form such a CPA system consists of a diffraction grating stretcher and compressor unit with the doped fiber as gain medium in between. The use of bulk optics and gratings require free space propagation and therefore need alignment. This diminishes the advantage of a fiber based laser system. In order to build a complete fiber based system, the grating stretcher can be replaced by fiber integrated optical devices such as a standard single-mode fiber [30] or a chirped fiber Bragg grating [31]. Both approaches allow stretching to the nanosecond regime and in particular chirped fiber Bragg gratings provide the adjustability of higher order dispersion. In the case of the compressor the situation becomes more delicate due to the high peak power obtained in the amplifier. Of course, chirped fiber Bragg gratings also offer the possibility of fiber integrated pulse compression. However, the pulse energy is typically restricted to the nanojoule range [32] because of enhanced nonlinear pulse distortion with increasing pulse peak power during the compression. Alternatively, chirped-periodic quasi-phasedmatched gratings [33] can be applied, offering scalability in peak power but are restricted in achievable time delay ( $\sim 50$  ps). Here we demonstrate an all fiber CPA system based on pulse compression in an air-guiding photonic band gap fiber which possesses anomalous dispersion and a significantly reduced nonlinearity due to the guiding of laser radiation in air. The experimental setup is shown in Fig. 16. Again, it starts with a passively mode-locked femtosecond Yb:KGW oscillator. The pulses are



**Fig. 16.** Experimental setup of the all fiber CPA system; OI: optical isolator, PC LMA: photonic crystal large-mode-area fiber

stretched in a 1.9 m long step-index single-mode fiber and are amplified using a 2.1 m long air-clad microstructured ytterbium-doped large-mode-area fiber. The 40 mW average seed power is amplified to a maximum of 8.2 W. A length of 2 m of the air-guiding photonic bandgap fiber is used to recompress the amplified, positively chirped pulses. The obtained 100-fs pulses with an average power as high as 6.0 W correspond to a peak power of 0.82 MW [34]. Nonlinear pulse distortions are not observed in the compressor fiber at this peak power level. Since such air-guiding fibers have approximately three orders of magnitude lower nonlinearity compared to conventional single-mode fibers, it indicates that this approach can be scaled to the  $\mu\text{J}$  pulse energy level with femtosecond pulse duration.

The potential of the air-core fiber to deliver these high peak power pulses has been demonstrated in the previous section. Due to the air-guiding fiber, damage will not limit the scalability. Furthermore, very high values of anomalous dispersion can be obtained especially on the long wavelength edge of the bandgap. Therefore in principle significantly longer initially stretched pulses can be recompressed with a short length of photonic crystal fiber. The main advantage of this approach is the possibility to build up a very compact and completely fiber integrated high peak power and high-energy fiber CPA system.

## 4 Conclusions and Outlook

In this contribution, the basic properties of photonic crystal waveguides and photonic bandgap fibers have been summarized and applications in ultrafast optics are demonstrated. Different photonic crystal waveguides have been discussed. By investigating wider waveguides relatively low propagation losses as small as 1.7 dB/mm could be found. Based on these waveguides, two devices were described in detail, filtering devices with up to 200 dB/mm suppression exploiting the occurrence of mini-stop bands and photonic crystal waveguide bends with efficiencies of up to 75% transmission per bend. By comparing photonic crystal waveguides with photonic crystal fibers the huge losses of the former become apparent. However one has to keep in mind that photonic crystal waveguides are aimed for the miniaturizing of integrated optics, where

sample lengths amounts to a few 100 $\mu\text{m}$  to mm. For such sizes propagation losses of a few dB/m or even dB/km are neither needed nor necessary.

The two-dimensional photonic bandgap also made a new type of fiber possible: the air-core photonic bandgap fiber. With its unique properties of low nonlinearity and anomalous dispersion we demonstrated the delivery of femtosecond pulses up to 3.3 MW peak power and showed flexible micro-machining as an application. Furthermore, we built an all fiber CPA system, where the air-core fiber replaced the grating compressor. It shows that compact and reliable all-fiber laser systems with megawatt peak-powers are possible. Although existing structures have made a big impact in the field of ultrafast optics, new devices will be investigated and applied in the near future. Surely they will be used to scale and improve existing ultrafast laser system. Also, new devices resulting from not yet fully exploited properties of photonic crystal structures seem to be feasible. One example is the realization of photonic crystal waveguides with ultra high dispersion. Calculations indicate dispersions up to  $\pm 10.000.000$  ps/nm/km. These high values occur at the band edge of band gap guided modes and also at the band edge of the index guided modes lying under the band gap. In contrast to band gap guided modes these modes are located under the light cone and thereby experience much lower propagation losses.

### Acknowledgements

This work is supported by the German Federal Ministry of Education and Research (BMBF, 01BK256 and 13N8336).

### References

1. J.D. Joannopoulos, R.D. Meade, J.N. Winn, *Photonic Crystals* (Princeton University Press, 1995).
2. E. Yablonovitch, Scientific American **285** (6), 46-54 (2001).
3. P. Vukusic, J.R. Sambles, Nature **424**, 852 (2003).
4. P. St. Russell, Science **299**, 358 (2003).
5. S. Noda, M. Imada, M. Okano, S. Ogawa, M. Mochizuki, A. Chutinan, IEEE J. Quant. Elect. **38** (7), 726 (2002).
6. K.M. Ho, C.T. Chan, C.M. Soukoulis, Phys. Rev. Lett. **65**, 3152 (1990).
7. S. Johnson, J.D. Joannopoulos, Opt. Express **8** (3), 173 (2001).
8. P. Lalanne, IEEE J. Quantum Elect. **38** (7), 800 (2002).
9. M. Mulot, S. Anand, M. Swillo, M. Qiu, B. Jaskorzynska, A. Talneau, J. Vac. Sci. Techn. B **21** (2), 900 (2003).
10. A. Talneau, L. Le Gouezigou, N. Bouadma, Opt. Lett. **26**, 1259 (2001).
11. S. Olivier, H. Benisty, C. Weisbuch, C.J.M. Smith, T.F. Krauss, R. Houdré, Opt. Express **11** (13), 1490 (2003).
12. S.J. McNab, N. Moll, Y.A. Vlasov, Opt. Express **11** (22), 2927 (2003)
13. M. Augustin, R. Iliew, H.-J. Fuchs, D. Schelle, C. Etrich, U. Peschel, S. Nolte, E.-B. Kley, F. Lederer, A. Tünnermann, Appl. Phys. Lett. **84**, 663 (2004).

14. M. Augustin, R. Iliew, H.-J. Fuchs, D. Schelle, C. Etrich, U. Peschel, S. Nolte, E.-B. Kley, F. Lederer, A. Tünnermann, Opt. Express **11** (24), 3284 (2003).
15. J.C. Knight, J. Broeng, T.A. Briks, P. St. J. Russell, Science **282**, 1476 (1998).
16. R.F. Cregan, B.J. Mangan, J.C. Knight, T.A. Birks, P.St.J. Russel, P.J. Roberts, and D.C. Allan, Science **285**, 1537 (1999).
17. J.C. Knight, Nature **424**, 847 (2003).
18. D. Ouzounov, F. Ahmad, A. Gaeta, M. Gallagher, K. Koch, D. Müller, and N. Venkataraman, "Dispersion and nonlinear propagation in air-core photonic bandgap fibers", in: *Conference on Lasers and Electro-Optics*, 2003 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 2003), paper CThV5.
19. D. Ouzounov, F. Ahmad, D. Müller, N. Venkataraman, M. Gallagher, K. Koch, A. Gaeta, Science **301**, 1702 (2003).
20. C. M. Smith, N. Venkataraman, M. T. Gallagher, D. Müller, J. A. West, N. F. Borrelli, D. C. Allan, K. W. Koch, Nature **424**, 657 (2003).
21. F. Benabid, J. C. Knight, G. Antonopoulos, P. St. J. Russell, Science **298**, 399 (2002).
22. F. Benabid, J. C. Knight, J. P. St. Russell, Opt. Express **10**, 1195 (2002).
23. J. Limpert, T. Schreiber, S. Nolte, H. Zellmer, T. Tunnermann, R. Iliew, F. Lederer, J. Broeng, G. Vienne, A. Petersson, and C. Jakobsen, Opt. Express **11**, 818 (2003).
24. H. Lubatschowski, G. Maatz, A. Heisterkamp, U. Hetzel, W. Drommer, H. Welling, W. Ertmer: Application of ultrashort laser pulses for intrastromal refractive surgery, Graefes. Arch. Clin. Exp. Ophthalmol. **238**, 33-39, 2000
25. see e.g. Fermann, Galvanauskas, Sucha, Ultrafast lasers: technology and applications, Decker, New York (Oct. 2002).
26. M. Will, J. Burghoff, J. Limpert, T. Schreiber, H. Zellmer, S. Nolte, A. Tünnermann, "Generation of photoinduced waveguides using a high repetition rate fiber CPA system", CLEO 2003/CWI6.
27. D. Strickland and G. Mourou, Opt. Commun. **56**, 219 (1985).
28. A. Galvanauskas, IEEE Journal on Selected Topics in Quantum Electronics **7**, 504 (2001).
29. J. Limpert, T. Clausnitzer, A. Liem, T. Schreiber, H.-J. Fuchs, H. Zellmer, E.-B. Kley, A. Tünnermann, Opt. Lett. **28** (20) 1984 (2003).
30. A. Liem, D. Nickel, J. Limpert, H. Zellmer, U. Griebner, S. Unger, A. Tünnermann, G. Korn, Appl. Phys. B **71**, 889 (2000).
31. A. Galvanauskas, M.E. Fermann, D. Harter, K. Sudgen, I. Bennion, Appl. Phys. Lett. **66**, 1053 (1995).
32. A. Galvanauskas, M.E. Fermann, D. Harter, J.D. Minelly, G.G. Vienne, J.E. Caplen, in: *Conference on Lasers and Electro-Optics*, Vol. 9 of 1996 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1996), pp. 495-496.
33. A. Galvanauskas, M.A. Arbore, M.M. Fejer, D. Harter, Opt. Lett. **23**, 1695 (1998).
34. J. Limpert, T. Schreiber, S. Nolte, H. Zellmer, A. Tünnermann, Opt. Express **11**, 3332 (2003).

## Part III

### Quantum Transport

# Novel Non-equilibrium Zero-resistance States in the High Mobility GaAs/AlGaAs System

R. G. Mani

Harvard University, Gordon McKay Laboratory of Applied Science  
9 Oxford Street, Cambridge, MA 02138 USA

**Abstract.** We illustrate various experimental features of the recently discovered radiation-induced zero-resistance states in the high mobility GaAs/AlGaAs system, with a special emphasis on the interplay between the radiation induced changes in the diagonal resistance and the Hall effect. Results examining the effect of simultaneous excitation at two radiation frequencies are also presented, along with reflection characteristics.

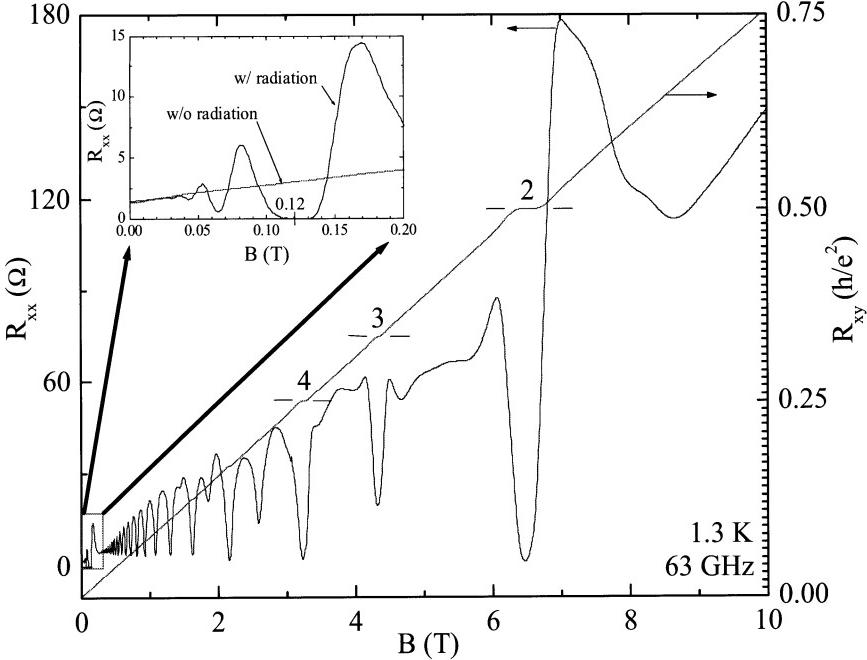
## 1 Introduction

The recent observation of novel radiation-induced zero-resistance states (ZRS) in the two-dimensional electron system has identified the possibility of remotely inducing vanishing resistance at low temperatures and weak magnetic fields [1]. This new effect [1,2,3] has provided a fresh impetus for the further study of 2D physics [4] because it suggests a new physical mechanism for suppressing the electrical resistance, in a sparsely studied non-equilibrium context [5,6,7,8,9,10,11,12,13,14,15,16,17,18].

As these novel zero-resistance states go together with an ordinary Hall effect, there is a need to understand the origin of vanishing resistance without Hall quantization in this context, unlike in the quantum Hall situation. From an experimental perspective, this calls for a further examination of the radiation induced changes in the Hall effect, and the study of the correlations between these modifications, and observed changes in the diagonal resistance [19]. Here, we examine this aspect, and exhibit also results illustrating the effect of simultaneous excitation at two radiation frequencies, along with reflection characteristics under microwave excitation.

## 2 Experiment

As indicated elsewhere [1,19], measurements were performed on Hall bars, square shaped devices, and Corbino rings, fabricated from GaAs/AlGaAs heterostructure devices, which were characterized, after a brief illumination by a red LED, by an electron density,  $n(4.2\text{ K}) \approx 3 \times 10^{11} \text{ cm}^{-2}$ , and an electron mobility up to  $\mu(1.5\text{ K}) \approx 1.5 \times 10^7 \text{ cm}^2/\text{Vs}$ . Lock-in based four-terminal electrical measurements were carried out with the sample mounted



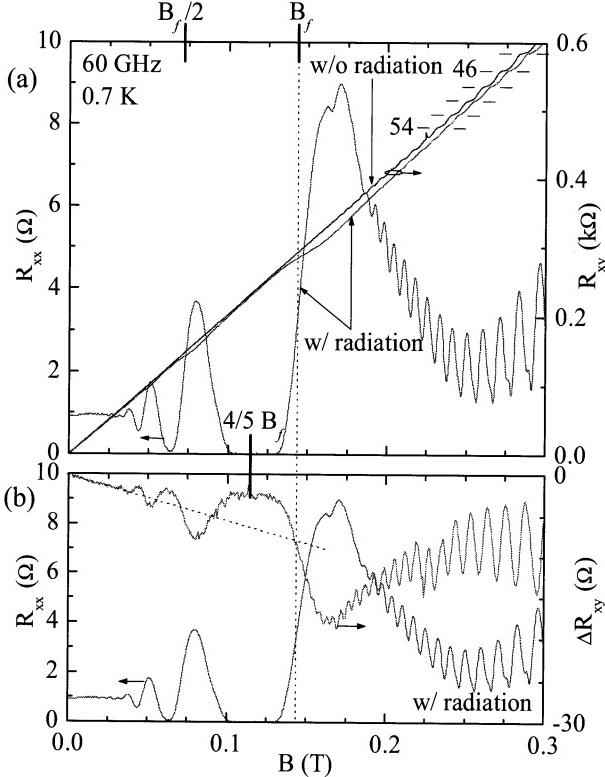
**Fig. 1.** (a): The figure shows the Hall ( $R_{xy}$ ) and diagonal ( $R_{xx}$ ) resistances of a GaAs/AlGaAs device under photo-excitation at 63 GHz. Quantum Hall effects (QHE) occur at high  $B$  as  $R_{xx}$  vanishes. (inset): An expanded view of the low- $B$  data. A radiation induced zero-resistance state is manifested about 0.12 Tesla

inside a waveguide and immersed in pumped liquid He-3 or He-4. Electromagnetic (EM) waves in the microwave part of the spectrum,  $3 \leq f \leq 170$  GHz, were generated using various tunable sources. For  $f \geq 27$  GHz, microwave radiation was coupled to the sample using the waveguide. At lower frequencies,  $3 \leq f \leq 20$  GHz, a coaxial line and a loop antenna served to channel the microwave radiation to the specimen.

### 3 Results

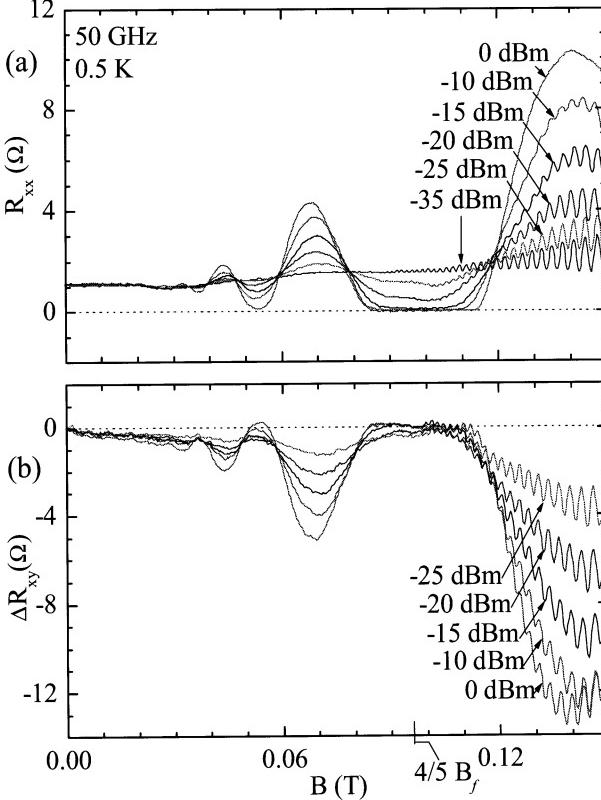
Figure 1 shows measurements, under 63 GHz microwave excitation, of the diagonal ( $R_{xx}$ ) and Hall ( $R_{xy}$ ) resistances to  $B = 10$  Tesla. For  $B \geq 0.2$  Tesla,  $R_{xx}$  and  $R_{xy}$  exhibit the usual quantum Hall behavior as the transport characteristics appear mostly insensitive to the microwave radiation [3,4]. The remarkable new effect occurs at low magnetic fields,  $B < 0.25$  Tesla in this instance, as shown in the inset of Fig. 1(a). Here, the resistance under microwave excitation falls below the dark resistance, leading to vanishing resistance over a broad magnetic field interval about 0.12 Tesla in Fig. 1 (inset) [1].

Figure 2 provides an expanded view of the low- $B$  transport under photo-excitation at 60 GHz. Fig. 2(a) shows a wide  $R_{xx}$  zero-resistance state about



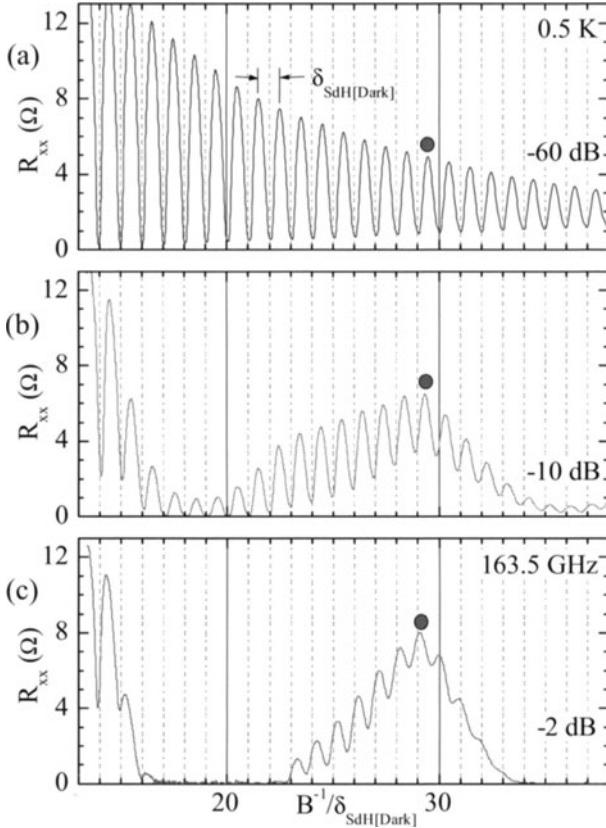
**Fig. 2.** (a): The figure exhibits the Hall resistance,  $R_{xy}$ , with (w/) and without (w/o) microwave radiation at  $f = 60$  GHz, in addition to  $R_{xx}$  under radiation. Here,  $R_{xx}$  vanishes about  $4/5B_f$ , although a plateau is not manifested in  $R_{xy}$  over the same  $B$  interval. Yet, there occur reductions in the magnitude of  $R_{xy}$  in the vicinity of the  $R_{xx}$  maxima. Note the remarkable shift of the quantized Hall plateaus, i.e., plateaus with  $R_{xy} = h/e^2$ , for  $46 \leq i \leq 54$ , to higher  $B$  under the influence of radiation. This feature could be understood as a consequence of a radiation induced change in the cross sectional area of the Fermi surface. (b): The radiation induced correction to the Hall resistance,  $\Delta R_{xy}$ , is shown along with  $R_{xx}$ . These  $\Delta R_{xy}$  oscillations correspond to a decrease in the magnitude of  $R_{xy}$ , while the slope in  $\Delta R_{xy}$  (dotted line) indicates a change in the slope of the  $R_{xy}$  curve under radiation. The latter feature, i.e., the change in slope, is consistent with the shift of the IQHE plateaus to higher  $B$  under radiation, see (a)

$4/5 B_f$ , and a close approach to vanishing resistance at the next lower- $B$  minimum, near  $4/9 B_f$ , which are the first two members of the series  $B = (4/[4j+1])B_f$ , with  $j = 1,2,3\dots$  Here,  $B_f = 2\pi fm^*/e$ ,  $m^*$  is an effective mass, and  $e$  is the charge. The remarkable feature in these novel zero-resistance states is the absence of simultaneous plateau formation in the Hall effect as in typical quantum Hall systems [4], which can be seen in the  $R_{xy}$  data that have been plotted vs. the right ordinate in Fig. 2(a). These Hall data show



**Fig. 3.** (a): The diagonal resistance  $R_{xx}$  is shown as a function of the magnetic field  $B$  for various source-radiation-intensities, in units of dBm, at 50 GHz. A zero-resistance state is observable about  $4/5 B_f$ . (b) This plot shows the radiation-induced change in the Hall resistance,  $\Delta R_{xy}$ , as a function of the magnetic field,  $B$ , with the radiation intensity as a parameter. The data indicate progressively stronger modulations in  $R_{xy}$  with increased radiation intensity. Note that  $\Delta R_{xy} \rightarrow 0$  over the  $B$  interval about  $4/5 B_f$  where  $R_{xx}$  vanishes in (a)

that, over the  $B$  interval corresponding with the  $4/5 B_f$  state, for example, the Hall resistance increases linearly vs. the magnetic field, and coincides with the  $R_{xy}$  that is observed without radiation. Yet, a careful comparison of the photo-excited (w/ radiation) Hall effect with the dark (w/o radiation) Hall effect reveals some peculiar changes upon irradiation. For example, at the  $R_{xx}$  maxima, there appear to be reductions in the magnitude of the irradiated  $R_{xy}$ . In addition, well above  $B_f$ , i.e.,  $B \geq 0.2$  Tesla,  $R_{xy}$  exhibits plateaus associated with the integral quantum Hall effect (IQHE), and plateaus of a given filling factor IQHE appear to shift to higher magnetic fields under the influence of radiation. Experiment showed, however, that this effect could be reversed by simply switching off the microwave excitation.

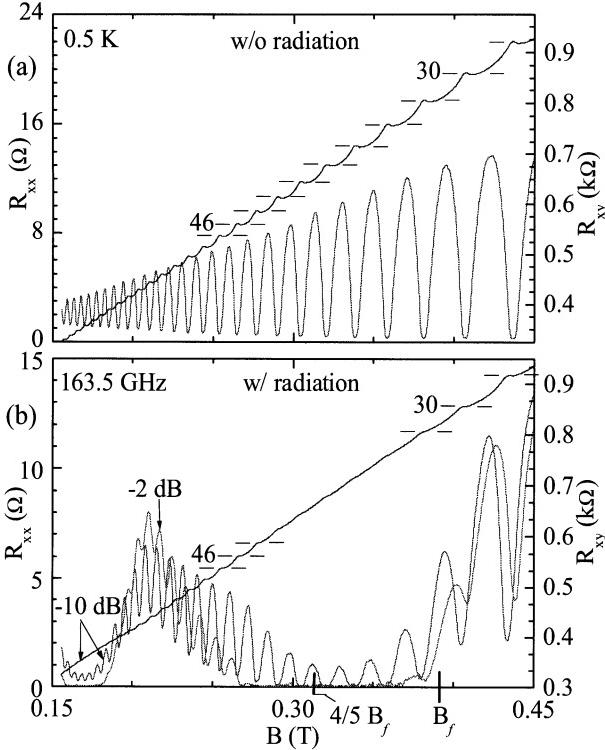


**Fig. 4.** This figure shows the evolution of the diagonal resistance  $R_{xx}$  under radiation, in the regime where strong Shubnikov-de Haas (SdH) oscillations are observable in  $R_{xx}$ . Top: This panel shows the SdH oscillations in the absence of radiation (-60 dB). The abscissa has been normalized by the period  $\delta_{SdH[Dark]}$  of these SdH oscillations. Note that SdH minima occur about integer values of the abscissa, implying  $R_{xx} \sim -\cos(B^{-1}/\delta_{SdH[Dark]})$ . Center: Here,  $R_{xx}$  under 163.5 GHz excitation has been exhibited with the radiation intensity attenuated to -10 dB. The radiation produces a strong modulation in the amplitude of the SdH oscillations, in conjunction with the radiation induced oscillatory magneto-resistance. Bottom:  $R_{xx}$  under photo-excitation, with the radiation intensity attenuated to -2 dB. Note the radiation induced zero-resistance states about  $17 \leq B^{-1}/\delta_{SdH[Dark]} \leq 23$  and  $34 < B^{-1}/\delta_{SdH[Dark]}$ , where the SdH oscillations have also vanished. This feature suggests a "re-entrant SdH effect" under photo-excitation. As the shaded disks mark a fixed filling factor in these panels, their shift along the abscissa is interpreted as a change in the cross sectional area of the Fermi surface under radiation

The radiation induced modification of the Hall effect is further examined in Fig. 2(b), where  $\Delta R_{xy}$  has been plotted along with  $R_{xx}$ . Here,  $\Delta R_{xy}$  has been obtained by subtracting the irradiated  $R_{xy}$  from the  $R_{xy}$  obtained in the dark. The key features in  $\Delta R_{xy}$  are: a) an oscillatory component that is period-commensurate with the  $R_{xx}$  oscillations, b) a finite slope in  $\Delta R_{xy}$  that is marked by the dotted line in the figure, and c) a  $\Delta R_{xy} \leq 0$  that represents a reduction in the magnitude of the Hall resistance under photo-excitation. Of these features, the reduction in slope of  $R_{xy}$  under photo-excitation appears to be consistent with the shift of fixed filling factor IQHE plateaus to higher- $B$  (Fig. 2(a)) under irradiation.

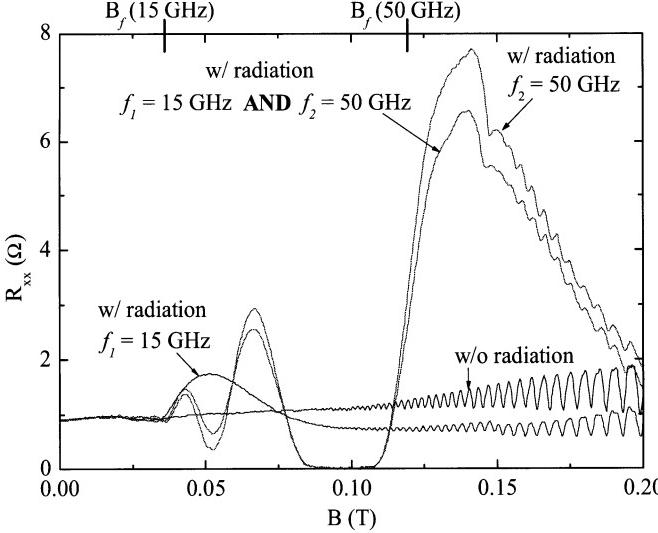
The radiation-intensity dependence of the  $R_{xx}$  oscillations and the associated radiation-induced changes in the Hall effect, i.e.,  $\Delta R_{xy}$ , are shown in Fig. 3(a) and Fig. 3(b), respectively, for 50 GHz excitation, with the source microwave intensity given in units of dBm. It is evident from this figure that increasing the radiation power produces progressively stronger variations in  $R_{xx}$  and  $\Delta R_{xy}$ . In particular, the minima in  $\Delta R_{xy}$  become deeper, as the maxima in  $R_{xx}$  grow stronger with increased power. Other noteworthy features include (i) the saturation of  $\Delta R_{xy}$  at zero-resistance in the vicinity of  $4/5 B_f$  as  $R_{xx} \rightarrow 0$ , and (ii)  $\Delta R_{xy} \rightarrow 0$  as  $B \rightarrow 0$ , even while  $R_{xx}$  saturates at a finite value as  $B \rightarrow 0$ . The latter feature is consistent with the finite slope observed in  $\Delta R_{xy}$  in Fig. 2(b).

Although not obvious in the above data, these radiation induced resistance oscillations can also occur over the same range of  $B$ , where strong SdH oscillations are exhibited by the specimen, over the precursor to the so-called quantum Hall regime [4]. That is, where the Landau level spacing exceeds  $k_B T$  and the level broadening. In order to illustrate this characteristic, data have been illustrated in Fig. 4, with microwave excitation of the specimen at  $f = 163.5$  GHz. Figure 4(a) shows that, in the dark, the specimen exhibits strong SdH oscillations over the range of  $B$  corresponding to filling factors  $26 \leq \nu \leq 76$ , i.e.,  $13 \leq B^{-1}/\delta_{\text{SdH}} \leq 38$ . Note that, over this range of  $B$ , spin degeneracy has not been lifted and, therefore, only even integral filling factor SdH minima are observable in the data. Upon increasing the radiation intensity, see Fig. 4(b) and (c), the radiation induced oscillations become manifested in  $R_{xx}$ , and therefore  $R_{xx}$  tends to vanish, for example, over a broad interval about  $B^{-1}/\delta_{\text{SdH}} = 20$  marking the  $4/5 B_f$  zero-resistance state. The zero-resistance effect in this regime is quite similar to the effect observed at lower  $f$ , where the Landau level broadening can be comparable to  $k_B T$ . Remarkably, as  $R_{xx} \rightarrow 0$  under the influence of radiation, the SdH oscillations also disappear. Yet, on either side of the  $4/5 B_f$  zero-resistance state, SdH oscillations remain observable. Indeed, such studies of the overlap of these two effects seem to suggest that the radiation induced zero-resistance state is perhaps tougher than the SdH zero-resistance state that is associated with the IQHE.



**Fig. 5.** This figure shows that the integer quantized Hall effect disappears under photo-excitation over the B-range where zero-resistance states are induced by microwave radiation. (a): In the absence of microwave excitation, the diagonal ( $R_{xx}$ ) and Hall ( $R_{xy}$ ) resistances in a GaAs/AlGaAs device exhibit strong Shubnikov-de Haas oscillations and integer quantum Hall effect (IQHE) respectively. Here, the even integer plateaus associated with the integer quantum Hall effect have been marked. (b) The Hall ( $R_{xy}$ ) and diagonal ( $R_{xx}$ ) resistances under microwave irradiation. Remarkably, the radiation induced zero-resistance state about, for example,  $4/5 B_f$ , suppresses the SdH oscillations and eliminates also the plateaus associated with the integer quantum Hall effect. Thus, for example, even integer  $i$  IQHE occur for  $R_{xy} = h/e^2$  with  $i \leq 32$ , and reappear for  $i > 44$ , while plateaus disappear between  $44 \geq i \geq 32$ . That is, microwave excitation produces a re-entrant behavior in the IQH- and SdH-effects

The interplay between the IQHE and the radiation induced ZRS is further examined in Fig. 5 through a direct comparison of the  $R_{xx}$  and  $R_{xy}$  under the without – (Fig. 5(a)) and with – (Fig 5(b)) microwave radiation conditions. Fig. 5(a) shows typical quantum Hall characteristics at high filling factors. Thus, one observes large amplitude SdH  $R_{xx}$  oscillations, where the minima approach zero-resistance, along with plateaus in  $R_{xy}$ , which appear at the expected quantized values  $h/ie^2$ . Although at this set-up limited measurement temperature,  $T = 0.5$  K, the SdH minima only approach zero-resistance,

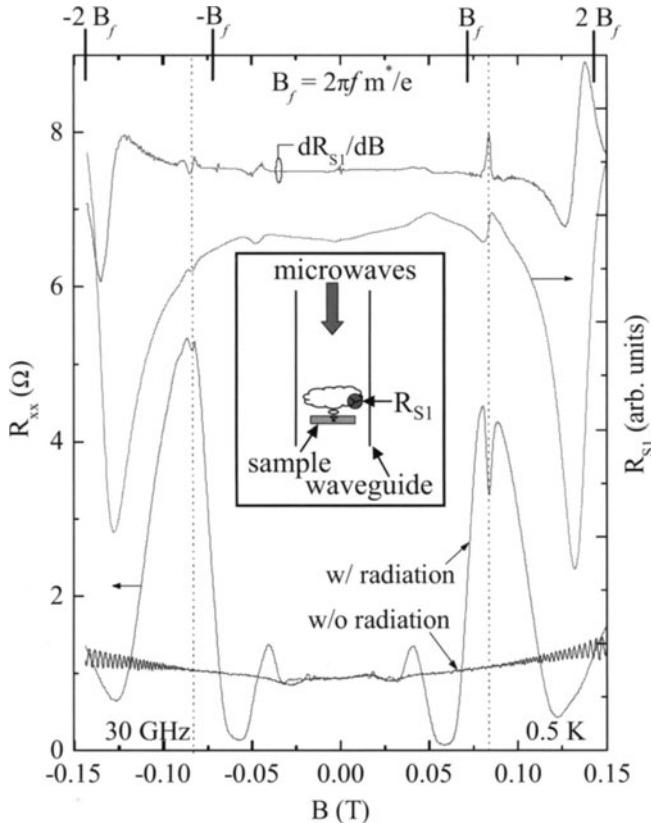


**Fig. 6.** This figure traces the effect of exciting the GaAs/AlGaAs device simultaneously at two different microwave frequencies. Thus, in addition to the dark (w/o radiation) magneto-resistance, there is exhibited the  $R_{xx}$  obtained under excitation at 50 GHz only, and at 15 GHz only. The associated field scales  $B_f$  are also shown in the figure. Finally, there is also exhibited the  $R_{xx}$  obtained under simultaneous excitation at 15 GHz and 50 GHz, which looks mostly like the trace obtained at 50 GHz. It could be that the 50 GHz radiation produces a "transparency" in the 2DES to the 15 GHz radiation [10]

a further reduction in  $T$  is likely to have produced resistance saturation of the SdH minima at  $R_{xx} = 0$ , even over this range of filling factors. Thus, measurements over this regime provide an indication of the effect that is to be observed, when the radiation induced zero-resistance states overlap with low-filling-factor quantum Hall effect and associated zero-resistance states.

Figure 5(b) exhibits the remarkable change in the  $R_{xx}$  and  $R_{xy}$  characteristics under the influence of radiation. Here, one observes that, as the amplitude of SdH oscillations in  $R_{xx}$  are reduced by microwave excitation in the vicinity of the microwave resistance minima, the plateaus and steps associated with the IQHE in  $R_{xy}$  also vanish. Indeed, in Fig. 5(b), a number of IQHE plateaus about the  $4/5 B_f$  zero-resistance state seem to have been replaced by an ordinary Hall effect. That is, under photo-excitation, the specimen seems to take on the characteristics, such as vanishing diagonal resistance and an ordinary Hall effect, of a perfect disorder free 2DES over the  $B$  interval which showed IQHE in the dark, see Fig. 5(b) [4].

It is also interesting to examine the effect of simultaneous excitation of the specimen at two different microwave frequencies, in light of theoretical encouragement for measurements along this direction [6,10]. Thus, in Fig. 6, we have shown measurements with excitation of the specimen at 15 GHz only,



**Fig. 7.** This figure examines the reflection/emission characteristics of the irradiated 2DES using a resistance-sensor above the sample. Thus, the figure exhibits the dark, and irradiated at 30 GHz,  $R_{xx}$ , along with the sensor  $R_{S1}$  response under photo-excitation. Also shown is the numerical derivative,  $dR_{S1}/dB$ , which serves to bring out weak features in the sensor signal. The  $R_{S1}$  signal shows two big features between  $B_f$  and  $2B_f$ , and the largest feature seems to originate from a two-photon process. There also occur additional weak features at lower  $B$ . The inset illustrates the measurement configuration

and at 50 GHz only. Here, at 50 GHz, a number of oscillations are induced by the radiation, and there occurs one broad zero-resistance state about  $4/5B_f$ . On the other hand, at 15 GHz, the radiation seems to produce an increase, followed by a decrease, in  $R_{xx}$ , with respect to the dark value above  $B_f$ . In addition to these data, the figure also shows the effect of simultaneous excitation at 15 GHz and 50 GHz. Remarkably, simultaneous excitation at two-frequencies results in a  $R_{xx}$  trace that looks very much like the trace at the higher  $f$ , i.e., 50 GHz, although there is a small decrease in the amplitude,

which could be attributed to a "breakdown" type heating effect as a result of combined excitation and increased radiation intensity at the specimen.

Finally, preliminary measurements examining possible reflection/emission characteristics of the high mobility 2DES under microwave excitation are illustrated in Fig. 7. For the reflection/emission experiments [see inset of Fig. 7], a resistance sensor was placed immediately above the sample, and its resistance was measured simultaneously along with the specimen characteristics, as radiation was applied from above.

Figure 7 exhibits the dark, and irradiated at 30 GHz,  $R_{xx}$ , along with the reflection/emission sensor  $R_{S1}$  response under photo-excitation. The irradiated  $R_{xx}$  shows giant oscillations and an approach to a zero-resistance state near  $4/5B_f$ . Note that the irradiated  $R_{xx}$  crosses the dark  $R_{xx}$  near  $B_f/j$ , with  $j = 1, 2, 3, \dots$  in addition to  $j = 1/2$ . Also shown in the figure is the numerical derivative,  $dR_{S1}/dB$ , which serves to bring out weak features in the sensor signal. The  $R_{S1}$  signal shows two significant features between  $\pm B_f$  and  $\pm 2B_f$ , and the largest feature might originate from a two-photon process. In the derivative signal, there is a very weak feature near  $-B_f$  (but not at  $+B_f$ ), which is attributed to electrically detected cyclotron resonance. The origin of other structure in the sensor signal is not fully understood.

In summary, we have reported the experimental characteristics of the irradiated high mobility 2DES in the regime of the novel radiation-induced zero-resistance states, with the emphasis on the interplay between the diagonal and off-diagonal resistances. The observed radiation induced modifications in off-diagonal resistance (Hall effect) are characterized by (a) a small reduction in the slope of the  $R_{xy}$  vs.  $B$  curve upon microwave excitation with respect to the dark value, which correlates with a change in the SdH period, (b) a reduction in the magnitude of the Hall resistance that goes together with an increase in the diagonal resistance  $R_{xx}$ . The radiation induced changes in  $R_{xx}$  and  $R_{xy}$  are comparable in magnitude, although  $\Delta R_{xy}$  is small compared to  $R_{xy}$ , and (c) a Hall resistance correction that disappears as the diagonal resistance vanishes. The radiation induced zero-resistance state also appears to overwhelm the SdH oscillations and the associated SdH zero-resistance state. Thus, as SdH oscillations disappear in the vicinity of  $B = 4/[(4j + 1)]B_f$  under photo-excitation so, remarkably, do the plateaus associated with the quantum Hall effect. We have also illustrated the effect of exciting the specimen at two different microwave frequencies, and exhibited some reflection characteristics.

### Acknowledgements

We acknowledge stimulating discussions with K. von Klitzing, V. Narayananamurti, J. H. Smet, and W. B. Johnson. The high quality MBE material was kindly provided by V. Umansky.

## References

1. R. G. Mani, J. H. Smet, K. von Klitzing, V. Narayanamurti, W. B. Johnson, and V. Umansky, *Nature* **420**, 646 (2002).
2. M. A. Zudov, R. R. Du, L. N. Pfeiffer and K. W. West, *Phys. Rev. Lett.* **90**, 046807 (2003).
3. R. Fitzgerald, *Phys. Today* **56** (4), 24 (2003).
4. R. E. Prange and S. M. Girvin, (eds.) *The Quantum Hall Effect*, 2nd Ed., Springer-Verlag (New York, 1990).
5. V. I. Ryzhii, *Sov. Phys. - Sol. St.* **11**, 2078 (1970).
6. J. C. Phillips, *Sol. St. Comm.* **127**, 233 (2003).
7. A. C. Durst, S. Sachdev, N. Read, and S. M. Girvin, *Phys. Rev. Lett.* **91**, 086803 (2003).
8. A. V. Andreev, I. L. Aleiner, and A. J. Millis, *Phys. Rev. Lett.* **91**, 056803 (2003).
9. J. Shi and X. C. Xie, *Phys. Rev. Lett.* **91**, 086801 (2003).
10. D. H. Lee and J. M. Leinaas, *Phys. Rev. B* **69**, 115336 (2004).
11. A. A. Koulakov and M. E. Raikh, *Phys. Rev. B* **68**, 115324 (2003).
12. V. Ryzhii and R. Suris, *J. Phys. Condens. Mat.* **15**, 6855 (2003).
13. X. L. Lei and S. Y. Liu, *Phys. Rev. Lett.*, **91**, 226805 (2003).
14. I. A. Dmitriev, A. D. Mirlin, and D. G. Polyakov, *Phys. Rev. Lett.* **91**, 226802 (2003).
15. I. G. Dmitriev et al., cond-mat/0310668 (unpublished).
16. K. Park, *Phys. Rev. B* **69**, 201301 (2004).
17. J. C. Phillips, cond-mat/0402112 (unpublished).
18. S. Fujita and K. Ito, cond-mat 0402174 (unpublished).
19. R. G. Mani et al., *Phys. Rev. Lett.* **92**, 146801 (2004); *Phys. Rev. B* **69**, 161306 (2004); *ibid.* **69**, 193304 (2004); R.G. Mani, *Physica E* **22**, 1 (2004).

# Oscillatory Photoconductivity of a Two-Dimensional Electron Gas in a Magnetic Field

I.A. Dmitriev<sup>1</sup>, M.G. Vavilov<sup>2</sup>, I.L. Aleiner<sup>3</sup>,  
A.D. Mirlin<sup>1,4</sup>, and D.G. Polyakov<sup>1\*</sup>

<sup>1</sup> Institut für Nanotechnologie, Forschungszentrum Karlsruhe  
76021 Karlsruhe, Germany

<sup>2</sup> Physics Dept., Massachusetts Institute of Technology  
Cambridge, MA 02139, USA

<sup>3</sup> Physics Dept., Columbia University  
New York, NY 10027, USA

<sup>4</sup> Institut für Theorie der Kondensierten Materie, Universität Karlsruhe  
76128 Karlsruhe, Germany

**Abstract.** We develop a theory of magneto-oscillations in the photoresistivity of a two-dimensional electron gas observed in recent experiments. It is shown that the effect is governed by a change of the electron distribution function induced by the microwave radiation. We analyze a nonlinearity with respect to both the *dc* field and the microwave power, as well as the temperature dependence determined by the inelastic relaxation rate.

## 1 Introduction

Recent experiments have discovered [1] that the resistivity of a high-mobility two-dimensional electron gas (2DEG) in GaAs/AlGaAs heterostructures subjected to microwave radiation of frequency  $\omega$  exhibits magneto-oscillations governed by the ratio  $\omega/\omega_c$ , where  $\omega_c$  is the cyclotron frequency. Subsequent experiments [2,3,4,5,6], working with very-high-mobility samples, yielded yet another discovery: for sufficiently high radiation power the minima of the oscillations evolve into “zero-resistance states” (ZRS), i.e., the dissipative resistance of a sample becomes vanishingly small.

These spectacular observations have attracted much theoretical interest. As was shown in Ref. [7], the ZRS can be understood as a direct consequence of the oscillatory photoconductivity (OPC), provided that the latter may become negative. A negative value of the OPC signifies an instability leading to the formation of spontaneous-current domains showing zero value of the observable resistance. Following this approach, a key issue which needs to be settled for understanding the data [1,2,3,4,5,6] is the identification of the microscopic mechanism of the OPC.

---

\* Corresponding author: polyakov@int.fzk.de

A mechanism of the OPC proposed in Ref. [8] is based on the effect of microwave radiation on electron scattering by impurities in a strong magnetic field (see also Ref. [9] for an earlier theory and Ref. [10] for a systematic theory). An alternative mechanism of the OPC was recently proposed in Ref. [11]. In contrast to Refs. [8,9,10], this mechanism is governed by a radiation-induced change of the electron distribution function. Because of the oscillations of the density of states (DOS),  $\nu(\varepsilon)$ , related to the Landau quantization, the correction to the distribution function acquires an oscillatory structure as well. This generates a contribution to the *dc* conductivity which oscillates with varying  $\omega/\omega_c$ . A distinctive feature of the contribution of Ref. [11] is that it is proportional to the inelastic relaxation time  $\tau_{\text{in}}$ . A comparison of the results of Refs. [10] and [11] shows that the latter contribution dominates if  $\tau_{\text{in}} \gg \tau_q$  (where  $\tau_q$  is the quantum, or single-particle, relaxation time due to impurity scattering), which is the case for the experimentally relevant temperatures.

The consideration of Ref. [11] is restricted to the regime which is linear in both the *ac* power and the *dc* electric field. The purpose of this paper is to develop a complete theory of the OPC governed by this mechanism, including nonlinear effects. We will demonstrate that the conductivity at a minimum becomes negative for a large microwave power and that a positive sign is restored for a strong *dc* bias, as it was assumed in Ref. [7]. Part of this work was presented in Ref. [12].

## 2 Non-equilibrium Distribution Function Induced by Microwave Radiation

We consider a 2DEG (mass  $m$ , density  $n_e$ , Fermi velocity  $v_F$ ) subjected to a transverse magnetic field  $B = (mc/e)\omega_c$ . We assume that the field is classically strong,  $\omega_c\tau_{\text{tr}} \gg 1$ , where  $\tau_{\text{tr}}$  is the transport relaxation time at  $B = 0$ . The photoconductivity  $\sigma_{\text{ph}}$  determines the longitudinal current flowing in response to a *dc* electric field  $\mathcal{E}_{\text{dc}}$ ,  $\mathbf{j} \cdot \mathbf{E}_{\text{dc}} = \sigma_{\text{ph}}\mathcal{E}_{\text{dc}}^2$ , in the presence of a microwave electric field  $\mathcal{E}_\omega \cos \omega t$ . The more frequently measured [1,2,3,5,6] longitudinal resistivity,  $\rho_{\text{ph}}$ , is given by  $\rho_{\text{ph}} \simeq \rho_{xy}^2 \sigma_{\text{ph}}$ , where  $\rho_{xy} \simeq eB/n_e c$  is the Hall resistivity, affected only weakly by the radiation.

We start with the formula for the *dc* conductivity,

$$\sigma_{\text{ph}} = 2 \int d\varepsilon \sigma_{\text{dc}}(\varepsilon) [-\partial_\varepsilon f(\varepsilon)] , \quad (1)$$

where  $f(\varepsilon)$  is the electron distribution function, and  $\sigma_{\text{dc}}(\varepsilon)$  determines the contribution of electrons with energy  $\varepsilon$  to the dissipative transport. In the leading approximation [10,11],  $\sigma_{\text{dc}}(\varepsilon) = \sigma_{\text{dc}}^{\text{D}} \tilde{\nu}^2(\varepsilon)$ , where  $\tilde{\nu}(\varepsilon) = \nu(\varepsilon)/\nu_0$  is the dimensionless DOS,  $\nu_0 = m/2\pi$  is the DOS per spin at zero  $B$  (we use  $\hbar = 1$ ), and  $\sigma_{\text{dc}}^{\text{D}} = e^2 \nu_0 v_F^2 / 2\omega_c^2 \tau_{\text{tr}}$  is the *dc* Drude conductivity per spin.

All interesting effects are due to a non-trivial energy dependence of the non-equilibrium distribution function  $f(\varepsilon)$ . The latter is found as a solution of the stationary kinetic equation

$$\begin{aligned} \mathcal{E}_\omega^2 \frac{\sigma^D(\omega)}{2\omega^2\nu_0} \sum_{\pm} \tilde{\nu}(\varepsilon \pm \omega) [f(\varepsilon \pm \omega) - f(\varepsilon)] \\ + \mathcal{E}_{dc}^2 \frac{\sigma_{dc}^D}{\nu_0 \tilde{\nu}(\varepsilon)} \frac{\partial}{\partial \varepsilon} \left[ \tilde{\nu}^2(\varepsilon) \frac{\partial}{\partial \varepsilon} f(\varepsilon) \right] = \frac{f(\varepsilon) - f_T(\varepsilon)}{\tau_{in}}, \end{aligned} \quad (2)$$

where the *ac* Drude conductivity per spin  $\sigma^D(\omega)$  is given by (we assume  $|\omega \pm \omega_c| \tau_{tr} \gg 1$ )

$$\sigma^D(\omega) = \sum_{\pm} \frac{e^2 \nu_0 v_F^2}{4\tau_{tr}(\omega \pm \omega_c)^2} \quad (3)$$

and  $\sigma_{dc}^D = \sigma^D(0)$ . On the right-hand side of Eq. (2), inelastic processes are included in the relaxation time approximation (more detailed discussion of the relaxation time  $\tau_{in}$  is relegated to the end of the paper), and  $f_T(\varepsilon)$  is the Fermi distribution. The left-hand side is due to the electron collisions with impurities in the presence of the external electric fields. The first term describes the absorption and emission of microwave quanta; the rate of these transitions was calculated in Ref. [11]. This term can be also extracted from the kinetic equation of Ref. [10]. The second term describes the effect of the *dc* field and can be obtained from the first one by taking the limit  $\omega \rightarrow 0$ .

Equation (2) suggests convenient dimensionless units for the strength of the *ac* and *dc* fields,

$$\mathcal{P}_\omega = \frac{\tau_{in}}{\tau_{tr}} \left( \frac{e\mathcal{E}_\omega v_F}{\omega} \right)^2 \frac{\omega_c^2 + \omega^2}{(\omega^2 - \omega_c^2)^2}, \quad (4)$$

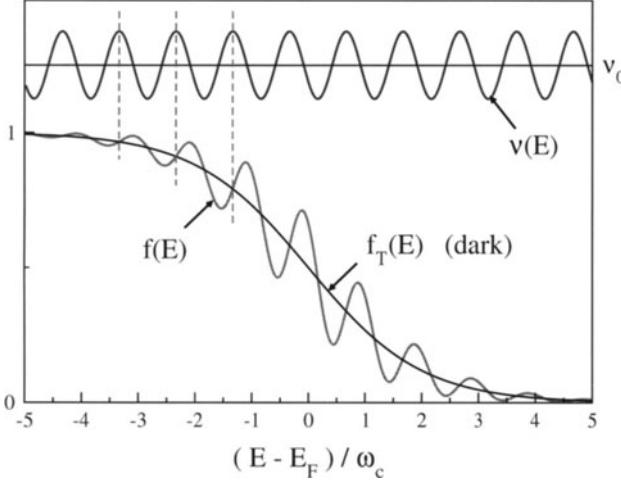
$$\mathcal{Q}_{dc} = \frac{2\tau_{in}}{\tau_{tr}} \left( \frac{e\mathcal{E}_{dc} v_F}{\omega_c} \right)^2 \left( \frac{\pi}{\omega_c} \right)^2. \quad (5)$$

Note that  $\mathcal{P}_\omega$  and  $\mathcal{Q}_{dc}$  are proportional to  $\tau_{in}$  and are infinite in the absence of inelastic relaxation processes.

We consider first the case of overlapping Landau levels (LLs), with the DOS given by  $\tilde{\nu} = 1 - 2\delta \cos \frac{2\pi\varepsilon}{\omega_c}$ , where  $\delta = \exp(-\pi/\omega_c \tau_q) \ll 1$ . Here  $\tau_q$  is the zero-*B* single-particle relaxation time, which is much shorter than the transport time in high-mobility structures,  $\tau_q \ll \tau_{tr}$  (because of the smooth character of a random potential of remote donors). The existence of a small parameter  $\delta$  simplifies solution of the kinetic equation (2). To first order in  $\delta$ , we look for a solution in the form

$$f(\varepsilon) = f_0(\varepsilon) + f_{osc}(\varepsilon) + O(\delta^2), \quad f_{osc}(\varepsilon) \equiv \delta \operatorname{Re} \left[ f_1(\varepsilon) e^{i \frac{2\pi\varepsilon}{\omega_c}} \right]. \quad (6)$$

We assume that the electric fields are not too strong [ $\mathcal{P}_\omega(\omega/T)^2 \ll 1$  and  $\mathcal{Q}_{dc}(\omega_c/T)^2 \ll 1$ ], so that the smooth part  $f_0(\varepsilon)$  is close to the Fermi distribution  $f_T(\varepsilon)$  at a bath temperature  $T \gg \omega_c$ ; otherwise, the temperature of



**Fig. 1.** Microwave-induced oscillatory distribution function  $f(\varepsilon)$  at  $T = \omega_c$

the electron gas is further increased due to heating. Smooth functions  $f_{0,1}(\varepsilon)$  change on a scale of the order of temperature. We obtain

$$f_{\text{osc}}(\varepsilon) = \delta \frac{\omega_c}{2\pi} \frac{\partial f_T}{\partial \varepsilon} \sin \frac{2\pi\varepsilon}{\omega_c} \frac{\mathcal{P}_\omega \frac{2\pi\omega}{\omega_c} \sin \frac{2\pi\omega}{\omega_c} + 4Q_{\text{dc}}}{1 + \mathcal{P}_\omega \sin^2 \frac{\pi\omega}{\omega_c} + Q_{\text{dc}}} . \quad (7)$$

Therefore, oscillations of the DOS  $\tilde{\nu}(\varepsilon)$  induce an oscillatory contribution  $f_{\text{osc}}(\varepsilon)$  to the distribution function, as illustrated in Fig. 1.

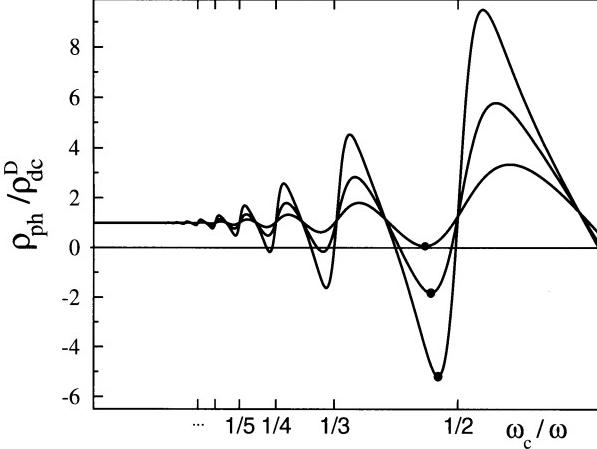
### 3 Oscillatory Photoconductivity

To calculate the photoconductivity, we substitute the result (7) for the distribution function into Eq. (1). Performing the energy integration in Eq. (1), we assume (in conformity with the experiment) that  $T$  is much larger than the Dingle temperature,  $T \gg 1/2\pi\tau_q$ . Under these conditions, the temperature smearing yields a dominant damping factor of the Shubnikov-de Haas oscillations,  $X/\sinh X$  with  $X = 2\pi^2 T/\omega_c$ . The terms of order  $\delta$  in Eq. (1) are then exponentially suppressed,

$$\delta \int \varepsilon \partial_\varepsilon f_T \cos \frac{2\pi\varepsilon}{\omega_c} \propto \delta \exp(-2\pi^2 T/\omega_c) \ll \delta^2 , \quad (8)$$

and can be neglected. The leading  $\omega$ -dependent contribution to  $\sigma_{\text{ph}}$  comes from the  $\delta^2$  term generated by the product of  $\partial_\varepsilon f_{\text{osc}}(\varepsilon) \propto \delta \cos \frac{2\pi\varepsilon}{\omega_c}$  and the oscillatory part  $-2\delta \cos \frac{2\pi\varepsilon}{\omega_c}$  of  $\tilde{\nu}(\varepsilon)$ . This term does survive the energy averaging,  $-\int d\varepsilon \partial_\varepsilon f_T \cos^2 \frac{2\pi\varepsilon}{\omega_c} \simeq 1/2$ . We thus find

$$\frac{\sigma_{\text{ph}}}{\sigma_{\text{dc}}^D} = 1 + 2\delta^2 \left( 1 - \frac{\mathcal{P}_\omega \frac{2\pi\omega}{\omega_c} \sin \frac{2\pi\omega}{\omega_c} + 4Q_{\text{dc}}}{1 + \mathcal{P}_\omega \sin^2 \frac{\pi\omega}{\omega_c} + Q_{\text{dc}}} \right) . \quad (9)$$



**Fig. 2.** Photoresistivity (normalized to the dark Drude value) for overlapping Landau levels vs  $\omega_c/\omega$  at fixed  $\omega\tau_q = 2\pi$ . The curves correspond to different levels of microwave power  $\mathcal{P}_\omega^{(0)} = \{0.24, 0.8, 2.4\}$ . Nonlinear  $I - V$  characteristics at the marked minima are shown in Fig. 3

Equation (9) is our central result. It describes the photoconductivity in the regime of overlapping LLs, including all non-linear (in  $\mathcal{E}_\omega$  and  $\mathcal{E}_{dc}$ ) effects. Let us analyze it in more detail. In the linear-response regime ( $\mathcal{E}_{dc} \rightarrow 0$ ) and for a not too strong microwave field, Eq. (9) yields a correction to the dark *dc* conductivity  $\sigma_{dc} = \sigma_{dc}^D(1 + 2\delta^2)$  which is linear in the microwave power:

$$\frac{\sigma_{ph} - \sigma_{dc}}{\sigma_{dc}} = -4\delta^2 \mathcal{P}_\omega \frac{\pi\omega}{\omega_c} \sin \frac{2\pi\omega}{\omega_c}, \quad (10)$$

in agreement with Ref. [11]. It is enlightening to compare Eq. (10) with the contribution of the effect of the *ac* field on the impurity scattering [8,9,10]. The analytic result, Eq. (6.11) of Ref. [10], in the notation of Eq. (2) is

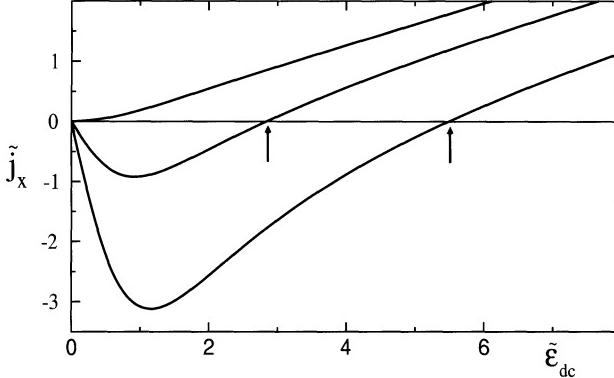
$$\frac{\sigma_{ph}^{[10]} - \sigma_{dc}}{\sigma_{dc}} = -12 \frac{\tau_q}{\tau_{in}} \delta^2 \mathcal{P}_\omega \left( \frac{\pi\omega}{\omega_c} \sin \frac{2\pi\omega}{\omega_c} + \sin^2 \frac{\pi\omega}{\omega_c} \right). \quad (11)$$

This result has a similar frequency dependence as Eq. (10); however, its amplitude is much smaller at  $\tau_{in} \gg \tau_q$ , i.e., the mechanism of Refs. [8,9,10] appears to be irrelevant. Physically, the effect of the *ac* field on the distribution function is dominant because it is accumulated during a diffusive process of duration  $\tau_{in}$ , whereas Refs. [8,9,10] consider only one scattering event.

With increasing microwave power, the photoconductivity saturates at the value

$$\frac{\sigma_{ph}}{\sigma_{dc}} = 1 - 8\delta^2 \frac{\pi\omega}{\omega_c} \cot \frac{\pi\omega}{\omega_c}, \quad \mathcal{P}_\omega \sin^2 \frac{\pi\omega}{\omega_c} \gg 1. \quad (12)$$

Note that although the correction is proportional to  $\delta^2 \ll 1$ , the factor  $8\pi(\omega/\omega_c) \cot(\pi\omega/\omega_c)$  is large in the vicinity of the cyclotron resonance har-



**Fig. 3.** Current–voltage characteristics [dimensionless current  $\tilde{j}_x = (\sigma_{\text{ph}}/\sigma_{\text{dc}}^D)\tilde{\mathcal{E}}_{\text{dc}}$  vs dimensionless field  $\tilde{\mathcal{E}}_{\text{dc}} = Q_{\text{dc}}^{1/2}$ ] at the points of minima marked by the circles in Fig. 1. The arrows show the *dc* field  $\tilde{\mathcal{E}}_{\text{dc}}^*$  in spontaneously formed domains

monics  $\omega = k\omega_c$  ( $k = 1, 2, \dots$ ), and allows the photo-induced correction to exceed in magnitude the dark conductivity  $\sigma_{\text{dc}}$ . In particular,  $\sigma_{\text{ph}}$  around minima becomes negative at  $\mathcal{P}_\omega > \mathcal{P}_\omega^* > 0$ , with the threshold value given according to Eq. (9) by

$$\mathcal{P}_\omega^* = \left( 4\delta^2 \frac{\pi\omega}{\omega_c} \sin \frac{2\pi\omega}{\omega_c} - \sin^2 \frac{\pi\omega}{\omega_c} \right)^{-1}. \quad (13)$$

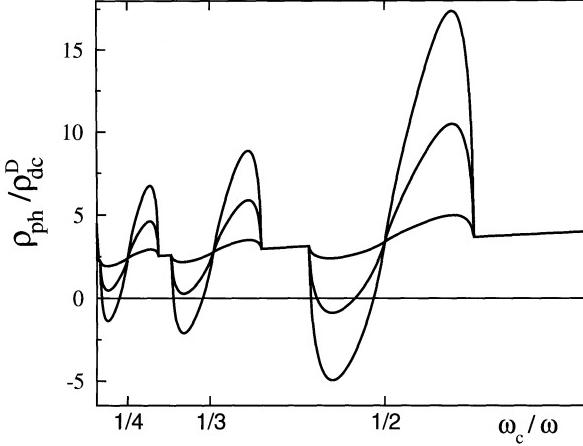
The evolution of a  $B$  dependence of the photoresistivity  $\rho_{\text{ph}}$  with increasing microwave power  $\mathcal{P}_\omega^{(0)} = \mathcal{P}_\omega(\omega_c = 0)$  is illustrated in Fig. 2.

Let us now fix  $\omega/\omega_c$  such that  $\mathcal{P}_\omega^* > 0$ , and consider the dependence of  $\sigma_{\text{ph}}$  on the *dc* field  $\mathcal{E}_{\text{dc}}$  at  $\mathcal{P}_\omega > \mathcal{P}_\omega^*$ . As follows from Eq. (9), in the limit of large  $\mathcal{E}_{\text{dc}}$  the conductivity is close to the Drude value and thus positive,  $\sigma_{\text{ph}} = (1 - 6\delta^2)\sigma_{\text{dc}}^D > 0$ . Therefore,  $\sigma_{\text{ph}}$  changes sign at a certain value  $\mathcal{E}_{\text{dc}}^*$  of the *dc* field, which is determined by the condition  $Q_{\text{dc}} = (\mathcal{P}_\omega - \mathcal{P}_\omega^*)/\mathcal{P}_\omega^*$ , see Fig. 3. The negative-conductivity state at  $\mathcal{E}_{\text{dc}} < \mathcal{E}_{\text{dc}}^*$  is unstable with respect to the formation of domains with a spontaneous electric field of the magnitude  $\mathcal{E}_{\text{dc}}^*$  [7].

Using Eqs. (9) and (2), we find the electric field  $\mathcal{E}_{\text{dc}}^*$  in the domains (measurable by local probe [6]),

$$e\mathcal{E}_{\text{dc}}^* = \frac{\omega_c}{\pi R_c} \left( \frac{\tau_{\text{tr}}}{2\tau_{\text{in}}} \right)^{1/2} \left[ \left( \frac{\mathcal{E}_\omega}{\mathcal{E}_\omega^*} \right)^2 - 1 \right]^{1/2}, \quad (14)$$

where  $\mathcal{E}_\omega^*$  is the threshold value of the *ac* field at which the zero-resistance state develops.



**Fig. 4.** Photoresistivity (normalized to the dark Drude value) for separated Landau levels vs  $\omega_c/\omega$  at fixed  $\omega\tau_q = 16\pi$ . The curves correspond to different levels of microwave power  $\mathcal{P}_\omega^{(0)} = \{0.004, 0.02, 0.04\}$

#### 4 Strong Magnetic Fields: Separated Landau Levels

We now turn to the regime of strong  $B$ ,  $\omega_c\tau_q/\pi \gg 1$ , where the LLs get separated. The DOS is then given (within the self-consistent Born approximation) by a sequence of semicircles of width  $2\Gamma = 2(2\omega_c/\pi\tau_q)^{1/2}$ :

$$\tilde{\nu}(\varepsilon) = \frac{2\omega_c}{\pi\Gamma^2} \sum_n \operatorname{Re} \sqrt{\Gamma^2 - (\varepsilon - n\omega_c - \omega_c/2)^2}. \quad (15)$$

We use Eqs. (1) and (2) to evaluate the OPC at  $\mathcal{Q}_{dc} \rightarrow 0$  to first order in  $\mathcal{P}_\omega$  and estimate the correction of the second order. We obtain

$$\frac{\sigma_{ph}}{\sigma_{dc}^D} = \frac{16\omega_c}{3\pi^2\Gamma} \left\{ 1 - \mathcal{P}_\omega \frac{\omega\omega_c}{\Gamma^2} \left[ \sum_n \Phi\left(\frac{\omega - n\omega_c}{\Gamma}\right) + O\left(\frac{\omega_c\mathcal{P}_\omega}{\Gamma}\right) \right] \right\}, \quad (16)$$

where

$$\Phi(x) = \frac{3x}{4\pi} \left[ \arccos(|x| - 1) - \frac{1 - |x|}{3} \sqrt{|x|(2 - |x|)} \right] \quad (17)$$

for  $|x| < 2$ , and  $\Phi(x) = 0$  otherwise. The photoresistivity for the case of separated LLs, Eq. (16), is shown in Fig. 4 for several values  $\mathcal{P}_\omega$  of the microwave power. Notice that a correction to Eq. (16) of second order in  $\mathcal{P}_\omega$  is still small even at  $\mathcal{P}_\omega > \mathcal{P}_\omega^* = \Gamma^2/\omega\omega_c$ , since  $\omega_c\mathcal{P}_\omega^*/\Gamma = \Gamma/\omega \ll 1$ . This means that it suffices to keep the linear-in- $\mathcal{P}_\omega$  term only even for the microwave power at which the linear-response resistance becomes negative.

As in the case of overlapping LLs, a negative value of the linear-response conductivity signals an instability leading to the formation of domains with

the field  $\mathcal{E}_{dc}^*$  at which  $\sigma_{ph}(\mathcal{E}_{dc}) = 0$ . It turns out, however, that for separated LLs the kinetic equation in the form of Eq. (2) yields zero (rather than expected positive) conductivity in the limit of strong  $\mathcal{E}_{dc}$ . This happens because elastic impurity scattering between LLs, inclined in a strong *dc* field, is not included in Eq. (2). The inter-LL transitions become efficient in *dc* fields as strong as  $\mathcal{E}_{dc}^* \simeq (\tau_{tr}/\tau_q)^{1/2} \omega_c^2 / ev_F$  [10], which actually gives the strength of the field in domains in the regime of separated Landau levels.

## 5 Temperature Dependence: Inelastic Relaxation Rate

Finally, we calculate the inelastic relaxation time  $\tau_{in}$ . Of particular importance is its  $T$  dependence which in turn determines that of  $\sigma_{ph}$ . At not too high  $T$ , the dominant mechanism of inelastic scattering is due to electron-electron (e-e) collisions. It is worth emphasizing that the e-e scattering does not yield relaxation of the total energy of the 2DEG and as such cannot establish a steady-state *dc* photoconductivity. That is to say the smearing of  $f_0(\varepsilon)$  in Eq. (6), which is a measure of the degree of heating, is governed by electron-phonon scattering. However, the e-e scattering at  $T \gg \omega_c$  does lead to relaxation of the oscillatory term  $f_{osc}$  [Eq. (7)] and thus determines the  $T$  behavior of the oscillatory contribution to  $\sigma_{ph}$ .

Quantitatively, the effect of electron-electron interaction is taken into account by replacing the right-hand side of Eq. (2) by  $-St_{ee}\{f\}$ , where the collision integral  $St_{ee}\{f\}$  is given by

$$St_{ee}\{f\} = \int d\varepsilon' \int dE A(E) \tilde{\nu}(\varepsilon_+) \tilde{\nu}(\varepsilon') \tilde{\nu}(\varepsilon'_-) \\ \times [-f(\varepsilon) f_h(\varepsilon_+) f(\varepsilon') f_h(\varepsilon'_-) + f_h(\varepsilon) f(\varepsilon_+) f_h(\varepsilon') f(\varepsilon'_-)] , \quad (18)$$

and  $f_h(\varepsilon) \equiv 1 - f(\varepsilon)$ ,  $\varepsilon_+ = \varepsilon + E$ ,  $\varepsilon'_- = \varepsilon' - E$ . The function  $A(E)$  describes the dependence of the matrix element of the screened Coulomb interaction on the transferred energy  $E$ ,

$$A(E) = \frac{1}{2\pi\epsilon_F} \ln \frac{\epsilon_F}{\max [E, \omega_c(\omega_c\tau_{tr})^{1/2}, \Gamma(\omega_c\tau_{tr})^{1/2}]} , \quad (19)$$

where  $\epsilon_F$  is the Fermi energy. Thus  $A(E)$  differs from the corresponding dependence for a clean 2DEG at zero  $B$  only by a change in the argument of the logarithm.

We linearize the collision integral and solve Eq. (2). For overlapping LLs, we put  $\tilde{\nu} = 1$  in accord with the accuracy of Eq. (9). Then only out-scattering processes contribute to the relaxation of the oscillatory part of the distribution function (7); the result is obtained by replacing  $\tau_{in} \rightarrow \tau_{ee}(\varepsilon, T)$  in Eq. (2) with [13]

$$\frac{1}{\tau_{ee}} = \frac{\pi^2 T^2 + \varepsilon^2}{4\pi\epsilon_F} \ln \frac{\epsilon_F}{\max [T, \omega_c(\omega_c\tau_{tr})^{1/2}]} . \quad (20)$$

We turn now to the case of separated LLs. In this case, due to oscillation of  $\tilde{\nu}$ , even the linearized collision integral gives rise to a non-trivial integral operator. Analytical solution of the kinetic equation with this collision operator does not seem feasible. However, up to a factor of order unity, we can replace the exact collision integral with the relaxation-time approximation, thus returning to Eq. (2) with

$$\frac{1}{\tau_{\text{in}}} \sim \frac{\omega_c}{\Gamma} \frac{T^2}{\epsilon_F} \ln \frac{\epsilon_F}{\max [T, \Gamma (\omega_c \tau_{\text{tr}})^{1/2}]} . \quad (21)$$

One sees that in both cases of overlapping and separated LLs the inelastic relaxation rate is proportional to  $T^2$ , so that the OPC  $\sigma_{\text{ph}} - \sigma_{\text{dc}}$  in the linear-in- $\mathcal{P}_\omega$  regime [Eqs. (10), (16)] scales as  $T^{-2}$ .

## 6 Summary and Comparison with Experiment

Our results are in overall agreement with the experimental findings [2,3]. The observed  $T$  dependence of the photoresistivity at maxima compares well with the predicted  $T^{-2}$  behavior. Typical parameters  $\omega/2\pi \simeq 50 - 100$  GHz,  $\tau_q \simeq 10$  ps yield  $\omega \tau_q/2\pi \simeq 0.5 - 1$  (overlapping LLs), and the experimental data indeed closely resemble Fig. 2. For  $T \sim 1$  K and  $\epsilon_F \sim 100$  K we find  $\tau_{\text{in}}^{-1} \sim 10$  mK, much less than  $\tau_q^{-1} \sim 1$  K, as assumed in our theory. Finally, for the microwave power  $\sim 1$  mW and the sample area  $\sim 1$  cm<sup>2</sup>, we estimate the dimensionless power  $\mathcal{P}_\omega^{(0)} \sim 0.005 - 0.1$ , which agrees with characteristic values for separated LLs (Fig. 4) but is noticeably less than the prediction for overlapping LLs (Fig. 2). The reason for this discrepancy remains to be clarified. Note also that the characteristic magnitude of the spontaneous field  $\mathcal{E}_{\text{dc}}^*$  in the domains, as follows from (14) is of order 1 V/cm, which compares well with local-probe measurements [6].

It is worth mentioning that the microwave-induced non-equilibrium distribution function should manifest itself also in thermodynamic quantities, such as the compressibility. It would be interesting to check this prediction experimentally.

## 7 Conclusions

To summarize, we have presented a theory of magneto-oscillations in the photoconductivity of a 2DEG. The parametrically largest contribution to the effect is governed by the microwave-induced change in the distribution function. We have analyzed the nonlinearity with respect to both the microwave and *dc* fields. The result takes an especially simple form in the regime of overlapping LLs, Eq. (9). For sufficiently strong microwave power, the linear-response resistivity at minima becomes negative, which implies a zero-resistance state by means of the formation of domains with spontaneous currents and electric

fields. We have further shown that the magnitude of the effect governed by the inelastic relaxation time increases as  $T^{-2}$  with lowering temperature.

As for future research, of central importance would be a detailed study (both experimental and theoretical) of transport through domains in the zero-resistance states. This should, in particular, elucidate the apparent activation temperature behavior at resistivity minima.

### Acknowledgements

We thank R.R. Du, K. von Klitzing, R.G. Mani, J.H. Smet, and M.A. Zudov for information about the experiments, and I.V. Gornyi for numerous stimulating discussions. This work was supported by the SPP “Quanten-Hall-Systeme” and SFB195 of the DFG, by NSF grants DMR02-37296 and EIA02-10376, by AFOSR grant F49620-01-1-0457, and by the RFBR.

## References

1. M.A. Zudov, R.R. Du, J.A. Simmons, J.L. Reno, Phys. Rev. B **64**, 201311(R) (2001); P.D. Ye, L.W. Engel, D.C. Tsui, J.A. Simmons, J.R. Wendt, G.A. Vawter, J.L. Reno, Appl. Phys. Lett. **79**, 2193 (2001).
2. R.G. Mani, J.H. Smet, K. von Klitzing, V. Narayanamurti, W.B. Johnson, V. Umansky, Nature **420**, 646 (2002); cond-mat/0303034, 0306388.
3. M.A. Zudov, R.R. Du, L.N. Pfeiffer, K.W. West, Phys. Rev. Lett. **90**, 046807 (2003).
4. C.L. Yang, M.A. Zudov, T.A. Knuutila, R.R. Du, L.N. Pfeiffer, K.W. West, Phys. Rev. Lett. **91**, 096803 (2003).
5. S.I. Dorozhkin, JETP Lett. **77**, 577 (2003).
6. R.L. Willett, L.N. Pfeiffer, K.W. West, cond-mat/0308406.
7. A.V. Andreev, I.L. Aleiner, A.J. Millis, Phys. Rev. Lett. **91**, 056803 (2003).
8. A.C. Durst, S. Sachdev, N. Read, S.M. Girvin, Phys. Rev. Lett. **91**, 086803 (2003).
9. V.I. Ryzhii, Sov. Phys. Solid State **11**, 2078 (1970); V.I. Ryzhii, R.A. Suris, B.S. Shchamkhalova, Sov. Phys. Semicond. **20**, 1299 (1986).
10. M.G. Vavilov, I.L. Aleiner, Phys. Rev. B **69**, 035303 (2004).
11. I.A. Dmitriev, A.D. Mirlin, D.G. Polyakov, Phys. Rev. Lett. **91**, 226802 (2003).
12. I.A. Dmitriev, M.G. Vavilov, I.L. Aleiner, A.D. Mirlin, D.G. Polyakov, cond-mat/0310668.
13. The energy  $\varepsilon$  is counted in Eq. (20) from the Fermi level. The  $\varepsilon$  dependence of  $\tau_{ee}^{-1}$  in Eq. (20) does not change the  $T^{-2}$  scaling [since the characteristic energy  $\varepsilon$  in Eq. (1) is of the order of  $T$ ], but only yields a numerical factor. In particular, the inelastic time entering Eq. (10) is  $\tau_{in} = 0.935 \tau_{ee}(\varepsilon = 0, T)$ .

# Molecular Wires in Electromagnetic Fields

Sigmund Kohler, Jörg Lehmann, Michael Strass, and Peter Hänggi

Institut für Physik, Universität Augsburg, 86135 Augsburg, Germany

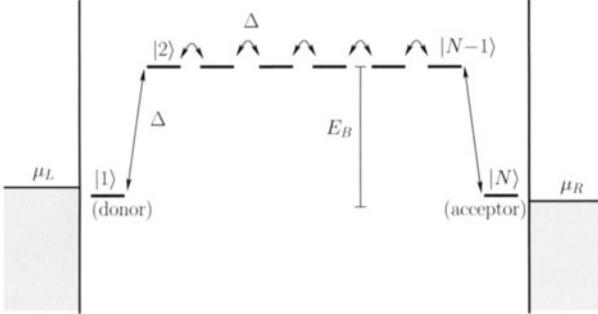
**Abstract.** We investigate the role of external electromagnetic fields on the conduction properties of bridged molecular wires. In particular, it is analyzed quantitatively how resonant excitations of electrons enhance the dc current and, simultaneously, lower the noise level of the current. The results from an exact numerical treatment are in good agreement with those obtained within an approximation scheme applicable at resonances.

Thirty years ago, Aviram and Ratner proposed in a seminal work [1] to build elements of electronic circuits—in their case a rectifier—with single molecules. In the present days their vision starts to become reality and the experimental and theoretical study of such systems enjoys a vivid activity [2–4]. Recent experimental progress has enabled reproducible measurements [5, 6] of weak tunneling currents through molecules which are coupled by chemisorbed thiol groups to the gold surface of external leads.

Typical energy scales in molecules are in the optical and the infrared regime, where basically all of the today's lasers operate. Hence, lasers represent a natural possibility to control atoms or molecules and also currents through them. It is for example possible to induce by the laser field an oscillating current in the molecule which under certain asymmetry conditions is rectified by the molecule. This results in a directed electron transport even in the absence of any applied voltage [7, 8]. Another theoretically predicted effect is the current suppression by the laser field [9, 10] which offers the possibility to control both the average current and the current noise.

Since the considered frequencies lie below typical plasma frequencies of metals, the laser light will be reflected at the metal surface, i.e., it does not penetrate the leads. Consequently, we assume that the leads' bulk properties are essentially unaffected by the laser field—in particular each lead remains close to equilibrium. Thus, it is sufficient to consider the influence of the driving solely in the molecule Hamiltonian. In addition, the energy of infrared light quanta is by far smaller than the work function of a common metal, which is of the order of 5 eV. This prevents the generation of a photo current, which otherwise would dominate the effects discussed below. For a quantitative description of an experiment, it might be necessary to take into account also the influence of the laser on the leads.

Most theoretical descriptions of the molecular conductivity in static situations are based on a scattering approach [11–13], or assume that the un-



**Fig. 1.** Bridged molecular wire consisting of  $N = 8$  sites of which the first and the last site are coupled to leads with chemical potentials  $\mu_L$  and  $\mu_R - eV$

derlying transport mechanism is an electron transfer reaction from the donor to the acceptor site and that the conductivity can be derived from the corresponding reaction rates [3].

Atoms and molecules in strong oscillating fields have been widely studied within a Floquet formalism [14, 15]. This suggests utilizing the tools that have been acquired in that area, thus, developing a transport formalism that combines Floquet theory for a driven molecule with the many-particle description of transport through a system that is coupled to ideal leads [8, 10, 16].

As an idealized model for the wire and the leads, we employ the so-called bridged molecule setup sketched in Fig. 1. This model has also been used to, e.g., investigate inelastic effects due to electron-vibrational coupling [17–20] and heat conduction [21]. A central issue in these works has been the length-dependence of the conduction thereby elucidating the underlying transport mechanisms [3]. Here, we address the influence of laser excitations on the length-dependence of the dc current and the low-frequency noise.

## 1 The Wire-Lead Model

wire-leads model of Fig. 1. It is convenient to separate the contributions from the molecule in the laser field, the ideal leads, and the molecule-leads coupling,

$$H(t) = H_{\text{molecule}}(t) + H_{\text{leads}} + H_{\text{molecule-leads}}. \quad (1)$$

The irradiated molecule is modeled within a tight-binding description taking into account  $N$  molecular orbitals  $|n\rangle$ , which are relevant for the transport. Disregarding the electron-electron interaction, the most general form of the Hamiltonian reads

$$H_{\text{molecule}}(t) = \sum_{n,n'} H_{nn'}(t) c_n^\dagger c_{n'}, \quad (2)$$

where the fermionic operators  $c_n$  and  $c_n^\dagger$  destroy and create, respectively, an electron in the molecular orbital  $|n\rangle$ . The sums extend over all tight-binding orbitals. The  $\mathcal{T}$ -periodic time-dependence of the single-particle Hamiltonian  $H_{nn'}(t) = H_{nn'}(t + \mathcal{T})$ , reflects the influence of the laser field with frequency  $\Omega = 2\pi/\mathcal{T}$ . As discussed above, we assume that the leads remain close to equilibrium and hence can be described by grand-canonical ensembles of electrons at temperature  $T$  and electro-chemical potential  $\mu_\ell$ ,  $\ell = L, R$ . Thus, the lead Hamiltonian reads

$$H_{\text{leads}} = \sum_{\ell q} \epsilon_q c_{\ell q}^\dagger c_{\ell q}, \quad (3)$$

where  $c_{\ell q}$  destroys an electron in state  $q$  in lead  $\ell$ . All expectation values of lead operators can be traced back to  $\langle c_{\ell q}^\dagger c_{\ell' q'} \rangle = \delta_{qq'} \delta_{\ell\ell'} f(\epsilon_q - \mu_\ell)$ , where  $f(\epsilon) = (1 + e^{\epsilon/k_B T})^{-1}$  denotes the Fermi function. The model is completed by the molecule-leads tunneling Hamiltonian

$$H_{\text{molecule-leads}} = \sum_q V_{Lq} c_{Lq}^\dagger c_1 + \sum_q V_{Rq} c_{Rq}^\dagger c_N + \text{h.c.}, \quad (4)$$

that connects the left (right) lead to the donor  $|1\rangle$  (acceptor  $|N\rangle$ ). Since we are not interested here in the effects that arise from the microscopic details of the molecule-lead coupling, we restrict our analysis in the following to energy-independent couplings, i.e.,  $\Gamma_\ell = 2\pi \sum_q |V_{\ell q}|^2 \delta(\epsilon - \epsilon_q) = \text{const.}$

## 2 Floquet Transport Theory

For the retarded Green function of the wire electrons, one finds after eliminating the leads the equation of motion [10]

$$[\mathcal{H}(t) - i\Sigma - i\hbar \frac{d}{dt}] G(t, t') = -\delta(t - t'), \quad (5)$$

where  $\mathcal{H}(t) = \sum_{n,n'} |n\rangle H_{nn'}(t) \langle n'|$  and  $2\Sigma = |1\rangle \Gamma_L \langle 1| + |N\rangle \Gamma_R \langle N|$  is the self-energy that results from the coupling to the leads. For the current defined as the change of the charge in the, *e.g.*, left lead,  $I_L = e(d/dt) \langle N_L \rangle$ , we find after some algebra that it assumes the commonly expected “scattering form” [13] but with periodically *time-dependent* transmission probabilities and, as well, an additional contribution that accounts for a  $\mathcal{T}$ -periodic charging/discharging of the wire [10, 16]. Only the former contributes to the time-averaged current

$$\bar{I} = \frac{e}{2\pi\hbar} \sum_{k=-\infty}^{\infty} \int d\epsilon \{ T_{LR}^{(k)}(\epsilon) f_R(\epsilon) - T_{RL}^{(k)}(\epsilon) f_L(\epsilon) \}, \quad (6)$$

where  $T_{LR}^{(k)}(\epsilon) = \Gamma_L \Gamma_R |G_{1N}^{(k)}(\epsilon)|^2$  is the transmission of an electron with energy  $\epsilon$  from the right lead to the left lead under the absorption (emission) of  $|k|$

photons if  $k > 0$  ( $k < 0$ ) and  $T_{RL}^{(k)}(\epsilon)$  accordingly.  $G_{1N}^{(k)}(\epsilon)$  denotes the relevant matrix elements of the Fourier transform of the retarded Green function

$$G^{(k)}(\epsilon) = \int_0^{\mathcal{T}} \frac{dt}{\mathcal{T}} e^{ik\Omega t} \int_{-\infty}^{+\infty} d\tau e^{i\epsilon\tau/\hbar} G(t, t - \tau). \quad (7)$$

Note that, consistent with Ref. [22], no ‘‘Pauli blocking factors’’  $1 - f_\ell$  appear in the current formula (6). In contrast to a static situation, this is of relevance here since for a driven system  $T_{LR}^{(k)}(\epsilon)$  and  $T_{RL}^{(k)}(\epsilon)$  are in general unrelated. Since the coefficients of the equation of motion (5) are  $\mathcal{T}$ -periodic, a complete solution can be constructed with the help of the Floquet ansatz  $|\psi_\alpha(t)\rangle = \exp[(-i\epsilon_\alpha/\hbar - \gamma_\alpha)t]|\Phi_\alpha(t)\rangle$ . The Floquet states  $|\Phi_\alpha(t)\rangle = \sum_k |\Phi_{\alpha k}\rangle \exp(-ik\Omega t)$  obey the time-periodicity of the Hamiltonian and fulfill in a Hilbert space that is extended by a periodic time coordinate the Floquet eigenvalue equation

$$[\mathcal{H}(t) - i\Sigma - i\hbar \frac{d}{dt}]|\Phi_\alpha(t)\rangle = (\epsilon_\alpha - i\hbar\gamma_\alpha)|\Phi_\alpha(t)\rangle. \quad (8)$$

Since the eigenvalue equation (8) is non-Hermitian, its eigenvalues  $\epsilon_\alpha - i\hbar\gamma_\alpha$  are generally complex valued and the (right) eigenvectors are not mutually orthogonal. Therefore, we need to solve also the adjoint Floquet equation yielding again the same eigenvalues but providing the adjoint eigenvectors  $|\Phi_\alpha^+(t)\rangle$ . Thus, we find for  $\tau > 0$  the retarded Green function

$$G(t, t - \tau) = -\frac{i}{\hbar} \sum_\alpha |\psi_\alpha(t)\rangle \langle \psi_\alpha^+(t - \tau)| \Theta(\tau) = G(t + \mathcal{T}, t + \mathcal{T} - \tau) \quad (9)$$

and, consequently,

$$G_{nn'}^{(k)}(\epsilon) = \sum_{\alpha, k'} \frac{\langle n | \Phi_{\alpha, k' + k} \rangle \langle \Phi_{\alpha, k'}^+ | n' \rangle}{\epsilon - (\epsilon_\alpha + k'\hbar\Omega - i\hbar\gamma_\alpha)}. \quad (10)$$

The current noise is given by the symmetrized auto-correlation function  $S_L(t, t') = \frac{1}{2} \langle \Delta I_L(t) \Delta I_L(t') + \Delta I_L(t') \Delta I_L(t) \rangle$  of the current fluctuation operator  $\Delta I_L(t) = I_L(t) - \langle I_L(t) \rangle$ . It can be shown that after the decay of all transients,  $S_L(t, t') = S_L(t + \mathcal{T}, t' + \mathcal{T})$  shares the time-periodicity of the driving. Therefore, it is possible to characterize the noise level by the time-averaged zero-frequency noise,  $\bar{S}_L = \int dt \int_0^{\mathcal{T}} dt S_L(t, t - \tau)/\mathcal{T}$ . Since the total charge is conserved, we find  $\bar{S}_L = \bar{S}_R = \bar{S}$ , where [10, 16]

$$\begin{aligned} \bar{S} = & \frac{e^2 \Gamma_L \Gamma_R}{2\pi\hbar} \sum_k \int d\epsilon \left\{ \Gamma_L \Gamma_R \left| \sum_{k'} G_{N1}^{(k'-k)}(\epsilon_k) [G_{N1}^{(k')}(\epsilon)]^* \right|^2 f_L(\epsilon) \bar{f}_L(\epsilon_k) \right. \\ & \left. + \left| \Gamma_L \sum_{k'} G_{1N}^{(k'-k)}(\epsilon_k) [G_{11}^{(k')}(\epsilon)]^* - iG_{1N}^{(-k)}(\epsilon_k) \right|^2 f_L(\epsilon) \bar{f}_R(\epsilon_k) \right\} \\ & + \text{same terms with the replacement } (L, 1) \leftrightarrow (R, N), \end{aligned} \quad (11)$$

with  $\bar{f}_{L/R} = 1 - f_{L/R}$  and  $\epsilon_k = \epsilon + k\hbar\Omega$ . In order to characterize the relative noise level, we employ the so-called Fano factor  $F = \bar{S}/e|\bar{I}|$  [23]. Expressions (6) and (11) contain as special cases prior findings: In the absence of any driving, the Floquet eigenvalues  $\epsilon_\alpha - i\hbar\gamma_\alpha$  reduce to the complex-valued eigenenergies; this implies  $G_{nn'}^{(k)} = 0$  for all  $k \neq 0$ , yielding the transmission probability for an electron with energy  $E$  of  $T(E) = \Gamma_L \Gamma_R |G_{N1}^{(0)}(E)|^2$ . Thus, the current and the noise in the static limit become

$$I_0 = \frac{e}{2\pi\hbar} \int dE T(E) [f_R(E) - f_L(E)], \quad (12)$$

$$S_0 = \frac{e^2}{2\pi\hbar} \int dE \left\{ T(E) [f_L(E)[1 - f_L(E)] + f_R(E)[1 - f_R(E)]] \right. \\ \left. + T(E)[1 - T(E)][f_R(E) - f_L(E)]^2 \right\}, \quad (13)$$

respectively [24]. In order to achieve for the noise an expression that depends only on the transmission probability  $T(E)$ , we have simplified the second line by use of the relation  $|\Gamma_L(\epsilon)G_{11}(\epsilon) + i|^2 = 1 - T(\epsilon)$  valid for undriven conductors [13]. Note that by contrast, in the time-dependent case, the noise expression (11) cannot be brought into such a convenient form and, thus in general, still depends on the phases of the propagator.

### 3 Conduction Properties of Driven Molecular Bridges

As a working model we consider a molecule consisting of a donor and an acceptor site and  $N - 2$  sites in between (cf. Fig. 1). Each of the  $N$  sites is coupled to its nearest neighbors by a hopping matrix element  $\Delta$ . The laser field renders each level oscillating in time with a position-dependent amplitude. The time-dependent molecule Hamiltonian is

$$H_{nn'}(t) = \Delta \sum_{n=1}^{N-1} (|n\rangle\langle n+1| + |n+1\rangle\langle n|) + E_1|1\rangle\langle 1| \\ + E_B \sum_{n=2}^{N-1} |n\rangle\langle n| + E_N|N\rangle\langle N| + A \cos(\Omega t) \sum_n x_n |n\rangle\langle n| \quad (14)$$

where  $x_n = (N + 1 - 2n)/2$  is the scaled position of site  $|n\rangle$ . The energy  $A$  equals the electron charge multiplied by the electrical field amplitude of the laser and the distance between two neighboring sites. The energies of the donor and the acceptor orbitals,  $|1\rangle$  and  $|N\rangle$ , are assumed to be at the level of the chemical potentials of the attached leads,  $E_1 = E_N \approx \mu_L \approx \mu_R$ . The bridge levels  $E_n$ ,  $n = 2, \dots, N - 1$ , lie  $E_B \gg \Delta$  above the chemical potential.

Below, we will evaluate the current and the noise for zero temperature and use a symmetric coupling,  $\Gamma_L = \Gamma_R = \Gamma$ . The hopping matrix element

$\Delta$  serves as the energy unit; in a realistic wire molecule,  $\Delta$  is of the order 0.1 eV. Thus, our chosen wire-lead hopping rate  $\Gamma = 0.1\Delta$  yields  $e\Gamma/\hbar = 2.56 \times 10^{-5}$  Ampère and  $\Omega \approx 10\Delta/\hbar$  corresponds to a laser frequency in the near infrared. For a typical distance of 5 Å between two neighboring sites, a driving amplitude  $A = \Delta$  is equivalent to an electrical field strength of  $2 \times 10^6$  V/cm.

Let us first discuss the static problem in the absence of the field, i.e. for  $A = 0$ . In the present case where the coupling between two neighboring sites is much weaker than the bridge energy,  $\Delta \ll E_B$ , one finds two types of eigenstates: One group of states is located on the bridge. It consists of  $N - 2$  levels with energies in the range  $[E_B - 2\Delta, E_B + 2\Delta]$ . In the absence of the driving field, these bridge states mediate the super-exchange between the donor and the acceptor. The two remaining states form a doublet whose states are approximately given by  $(|1\rangle \pm |N\rangle)/\sqrt{2}$ . Its splitting can be estimated in a perturbational approach [25] and is approximately given by  $2\Delta(\Delta/E_B)^{N-2}$ . Thus, the wire can be reduced to a two-level system with the effective tunnel matrix element  $\Delta_{DA} = \Delta \exp(-\kappa(N-2))$ , where  $\kappa = \ln(E_B/\Delta)$ . If the chemical potentials of the leads are such that  $\mu_L > E_D$  and  $\mu_R < E_A$ , i.e., for a sufficiently large voltage, the current (12) is dominated by the total transmission and for  $\Gamma \gg \Delta_{DA}$  can be evaluated to read

$$I_0 = \frac{2e|\Delta|^2}{\Gamma} e^{-2\kappa(N-2)}. \quad (15)$$

For the explicit calculation see, e.g., Ref. [26]. In particular, one finds an exponentially decaying length dependence of the current [3, 12]. Moreover, in this limit, it is also possible to evaluate explicitly the zero-frequency noise to obtain the Fano factor  $F = \bar{S}/e|\bar{I}| = 1$ . This value has a direct physical interpretation: Since the transmissions of electrons across a large barrier are “rare and uncorrelated events”, they obey Poisson statistics and, thus, variance and mean value are equal which translates to a Fano factor of one [23].

### 3.1 Resonant Excitations

The magnitude of the current changes significantly when a driving field with a frequency  $\Omega \approx E_B/\hbar$  is switched on. The resonant bridge levels merge with the donor and acceptor states to form a Floquet state. This opens a direct transport channel resulting in an enhancement of the electron current.

In order to estimate the magnitude of the current through the resonantly driven wire, we disregard all bridge levels besides the one that is in resonance with the donor and the acceptor. Let us assume that this resonant bridge level  $|\psi_B\rangle$  extends over the whole bridge such that it occupies the sites  $|2\rangle, \dots, |N-1\rangle$  with equal probability  $1/\sqrt{N-2}$ . Accordingly, the overlap between the bridge level and the donor/acceptor becomes

$$\langle 1 | H_{\text{molecule}} | \psi_B \rangle = \frac{\langle 1 | H_{\text{molecule}} | 2 \rangle}{\sqrt{N-2}} = \frac{\Delta}{\sqrt{N-2}} \quad (16)$$

and  $\langle \psi_B | H_{\text{molecule}} | N \rangle$  accordingly, while the resonance condition defines the energy of the bridge level as

$$\langle \psi_B | H_{\text{molecule}} | \psi_B \rangle = \hbar\Omega \quad (17)$$

(recall that we have assumed  $E_D = E_A = 0$ ).

We apply an approximation in the spirit of the one described in [26] and derive a *static* effective Hamiltonian that describes the *time-dependent* system. We start by transforming with the unitary operator

$$S(t) = \exp \left\{ -i \sum_{n=2}^{N-1} |n\rangle \langle n| \Omega t - i \frac{A}{\hbar\Omega} \sum_{n=1}^N |n\rangle \langle n| \sin(\Omega t) \right\}. \quad (18)$$

Note that  $S(t)$  obeys the  $T$ -periodicity of the original wire Hamiltonian (14). As a consequence, the transformed wire Hamiltonian

$$\tilde{H}_{\text{molecule}}(t) = S^\dagger(t) H_{\text{molecule}}(t) S(t) - i\hbar S^\dagger(t) \dot{S}(t) \quad (19)$$

is  $T$ -periodic as well. For  $\hbar\Omega \gg \Delta$ , we can separate time-scales and average  $\tilde{H}_{\text{molecule}}(t)$  over the driving period. In the subspace spanned by  $|1\rangle$ ,  $|\psi_B\rangle$ , and  $|N\rangle$ , the time-averaged wire Hamiltonian reads

$$H_{\text{molecule,eff}} = \int_0^T \frac{dt}{T} \tilde{H}_{\text{molecule}}(t) = \begin{pmatrix} 0 & b & 0 \\ b & 0 & b \\ 0 & b & 0 \end{pmatrix}, \quad (20)$$

with the effective tunnel matrix element

$$b = \frac{J_1(A/\hbar\Omega)}{\sqrt{N-2}} \Delta, \quad (21)$$

and  $J_1$  the first-order Bessel function of the first kind.

The situation described by the Hamiltonian (20) is essentially the following: The central site  $|\psi_B\rangle$  is coupled by matrix elements  $b$  to the donor and the acceptor site. Since the latter in turn couple to the external leads with a self energy  $\Gamma/2$ , their density of states is

$$\rho(E) = \frac{1}{\pi} \frac{\Gamma/2}{E^2 + (\Gamma/4)^2}. \quad (22)$$

Then, the tunneling of the electrons from and to the central site is essentially given by the golden rule rate  $w = 2\pi|b|^2\rho(0)/\hbar$ . Like in the static case, we assume that the chemical potential of the left (right) lead lies above (below) the on-site energy of the donor (acceptor) and that therefore the donor is always occupied while the acceptor is always empty. Then, the rate of electrons tunneling from the central site to the acceptor is given by the golden rule rate times the occupation probability  $p$  of the state  $|\psi_B\rangle$ . Accordingly,

the rate of electron from the donor to  $|\psi_B\rangle$  is given by  $w$  times the probability  $(1 - p)$  to find the central site empty. Consequently, the occupation of the resonant bridge level evolves according to  $\dot{p} = w(1 - p) - wp$ . This equation has the stationary solution  $p = 1/2$  and, thus, for resonant excitations, the dc contribution of the time-dependent current is given by

$$\bar{I}_{\text{res}} = e w p = e \frac{2A^2 \Delta^2}{(N - 2)\hbar^3 \Omega^2 \Gamma}. \quad (23)$$

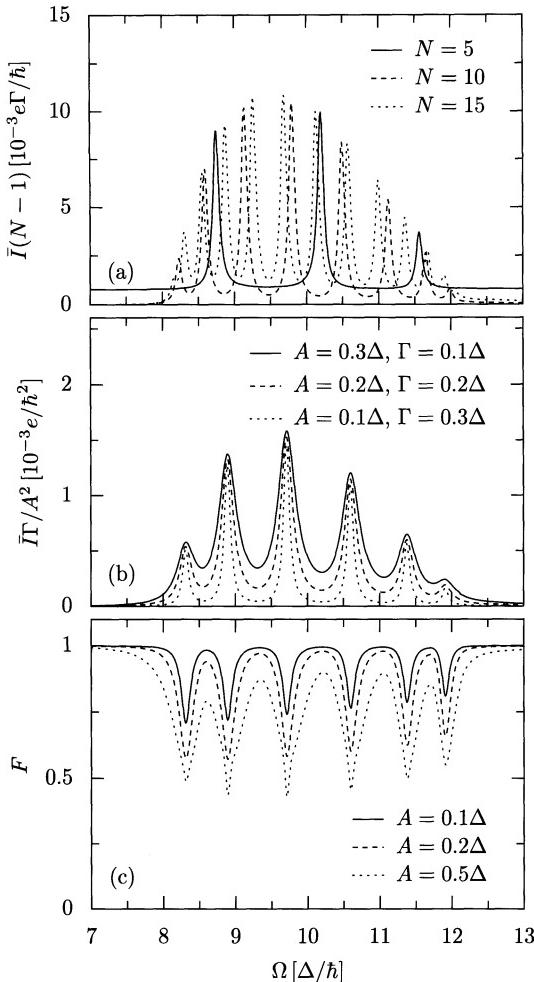
Here, we have used for small arguments of the Bessel function the approximation  $J_1(x) \approx x$ . The dc current (23) obeys an intriguing scaling behavior as a function of the wire length: Instead of the exponentially decaying length dependence (15) that has been found for the static case, in the presence of resonant driving, a scaling  $\bar{I} \propto 1/N$  emerges. In particular for longer wires, this means that the external field enhances the conductance significantly.

### 3.2 Numerical Results

In order to corroborate the analytical estimates presented above, we treat the transport problem defined by the wire Hamiltonian (14) numerically by solving the corresponding Floquet equation (8) and a subsequent evaluation of the expressions (6) and (11) for the dc current and the zero-frequency noise, respectively. For a wire with  $N = 5$  sites, one finds peaks in the current when the driving frequency matches the energy difference between the donor/acceptor doublet and one of the  $N - 2 = 3$  bridge levels, cf. the solid line in Fig. 2a. The applied voltage is always chosen so small that the bridge levels lie below the chemical potentials of the leads. In Fig. 2a the scale of the abscissa is chosen proportional to  $(N - 1)$  such that it suggests a common envelope function. Furthermore, we find from Fig. 2b that the dc current is proportional to  $A^2/\Gamma$  provided that  $A$  is sufficiently small and  $\Gamma$  sufficiently large. The numerical results indicate that the height of the current peaks obeys  $\bar{I}_{\text{peak}} \propto \frac{A^2}{(N-1)\Gamma}$  which is essentially in accordance with our analytical estimate (23). The main discrepancy comes from the fact that the overlap between the resonant level and the donor/acceptor differs from the estimate (16) by a numerical factor of the order one. Moreover, Fig. 2c demonstrates that at the resonances, the Fano factor assumes values considerably lower than one as expected for the transport through a resonant single level [24].

## 4 Conclusions

We have outlined the Floquet transport formalism which has been derived originally in Refs. [10, 16] to provide explicit expressions for the current through time-dependent nanoscale conductors and its zero-frequency noise. With this formalism, the conductance properties of bridged molecular wires



**Fig. 2.** Exact numerical solution within the Floquet formalism. (a) Average current  $\bar{I}$  as a function of the driving frequency  $\Omega$  for various wire length  $N$ . The scaled amplitude is  $A = 0.1\Delta$ ; the applied voltage  $\mu_R - \mu_L = 5\Delta/e$ . The other parameters read  $\Gamma = 0.1\Delta$ ,  $E_B = 1\Delta$ , and  $k_B T = 0$ . (b) Average current for various driving amplitudes  $A$  and coupling strengths  $\Gamma$  for a wire of length  $N = 8$ . (c) Fano factor  $F = \bar{S}/e\bar{I}$  for the wire length  $N = 8$  and the wire-lead coupling  $\Gamma = 0.1\Delta$

have been investigated. This has revealed that resonant excitations from the levels that connect the molecule to the external leads to bridge levels yield peaks in the current as a function of the driving frequency. In a regime with weak driving and weak electron-lead coupling,  $\Delta \gg \Gamma, A$ , the peak heights scale with the coupling strength, the driving amplitude, and the wire length. The laser irradiation induces a large current enhancement of several orders of magnitude and also can reduce the current noise level. The observation of

these resonances could serve as an experimental starting point for the more challenging attempt of measuring quantum ratchet effects [7, 8] or current switching by laser fields [9, 10].

### Acknowledgements

We acknowledge helpful discussions with Sébastien Camalet and Gert-Ludwig Ingold. This work has been supported by the Volkswagen-Stiftung under Grant No. I/77 217 and the DFG Sonderforschungsbereich 486.

## References

1. A. Aviram and M. A. Ratner, *Chem. Phys. Lett.* **29**, 277 (1974).
2. C. Joachim, J. K. Gimzewski, and A. Aviram, *Nature* **408**, 541 (2000).
3. A. Nitzan, *Annu. Rev. Phys. Chem.* **52**, 681 (2001).
4. *Special Issue: Processes in Molecular Wires*, Vol. 281 of *Chem. Phys.*, edited by P. Hänggi, M. Ratner, and S. Yaliraki (Elsevier, Amsterdam, 2002), pp. 111–502.
5. X. D. Cui, A. Primak, X. Zarate, J. Tomfohr, O. F. Sankey, A. L. Moore, T. A. Moore, D. Gust, G. Harris, and S. M. Lindsay, *Science* **294**, 571 (2001).
6. J. Reichert, R. Ochs, D. Beckmann, H. B. Weber, M. Mayor, and H. von Löhneysen, *Phys. Rev. Lett.* **88**, 176804 (2002).
7. J. Lehmann, S. Kohler, P. Hänggi, and A. Nitzan, *Phys. Rev. Lett.* **88**, 228305 (2002).
8. J. Lehmann, S. Kohler, P. Hänggi, and A. Nitzan, *J. Chem. Phys.* **118**, 3283 (2003).
9. J. Lehmann, S. Camalet, S. Kohler, and P. Hänggi, *Chem. Phys. Lett.* **368**, 282 (2003).
10. S. Camalet, J. Lehmann, S. Kohler, and P. Hänggi, *Phys. Rev. Lett.* **90**, 210602 (2003).
11. Y. Meir and N. S. Wingreen, *Phys. Rev. Lett.* **68**, 2512 (1992).
12. V. Mujica, M. Kemp, and M. A. Ratner, *J. Chem. Phys.* **101**, 6849 (1994).
13. S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1995).
14. N. L. Manakov, V. D. Ovsiannikov, and L. P. Rapoport, *Phys. Rep.* **141**, 319 (1986).
15. M. Grifoni and P. Hänggi, *Phys. Rep.* **304**, 229 (1998).
16. S. Camalet, S. Kohler, and P. Hänggi, cond-mat/0402182 (2004).
17. W. B. Davis, M. R. Wasielewski, M. A. Ratner, V. Mujica, and A. Nitzan, *J. Phys. Chem.* **101**, 6158 (1997).
18. D. Segal, A. Nitzan, W. B. Davis, M. R. Wasielewski, and M. A. Ratner, *J. Phys. Chem.* **104**, 3817 (2000).
19. D. Segal and A. Nitzan, *Chem. Phys.* **281**, 235 (2002).
20. V. May, *Phys. Rev. B* **66**, 245411 (2002).
21. D. Segal, A. Nitzan, and P. Hänggi, *J. Chem. Phys.* **119**, 6840 (2003).
22. S. Datta and M. P. Anantram, *Phys. Rev. B* **45**, 13761 (1992).
23. U. Fano, *Phys. Rev.* **72**, 26 (1947).

24. Ya. M. Blanter and M. Büttiker, Phys. Rep. **336**, 1 (2000).
25. M. A. Ratner, J. Phys. Chem. **94**, 4877 (1990).
26. S. Kohler, S. Camalet, M. Strass, J. Lehmann, G.-L. Ingold, and P. Hänggi, Chem. Phys. **296**, 243 (2004).

# Coulomb Drag as a Probe of Quantum Hall Systems

Felix von Oppen<sup>1,2</sup>, I.V. Gornyi<sup>3\*</sup>, and A.D. Mirlin<sup>3,4\*\*</sup>

<sup>1</sup> Institut für Theoretische Physik, Freie Universität Berlin  
14195 Berlin, Germany

<sup>2</sup> Dept. of Condensed Matter Physics  
Weizmann Institute of Science, Rehovot, Israel

<sup>3</sup> Institut für Nanotechnologie  
Forschungszentrum Karlsruhe, 76021 Karlsruhe, Germany

<sup>4</sup> Institut für Theorie der Kondensierten Materie  
Universität Karlsruhe, 76128 Karlsruhe, Germany

**Abstract.** Experiments on Coulomb drag in double-layer systems in the quantum Hall regime have observed a number of surprises which were only partially understood for some time. The most striking observations are that Coulomb drag can become negative in the regime of high Landau levels and that its temperature dependence is non-monotonous. In this contribution, we review a theory of Coulomb drag in high Landau levels for the ballistic regime which is in good agreement with the experiments. Our starting point is the diagrammatic approach to drag, treating the interlayer interaction perturbatively and accounting for disorder within the self-consistent Born approximation. Our theory shows that drag in high Landau levels is an interplay of two contributions arising from different sources for particle-hole asymmetry, namely the curvature of the zero-field electron dispersion and the Landau-level density of states.

Frictional drag [1,2] between two parallel two-dimensional electron systems has become a powerful tool to study quantum Hall systems. When a current  $I$  is applied to one (active) layer, interlayer interactions will impart momentum to the carriers of the second, passive layer. If no current is allowed to flow in the passive layer, a voltage  $V$  builds up there, neutralizing the momentum transfer. The ratio  $V/I$  is known as drag resistance or transresistance. Depending on the strength of the interlayer interaction – tunable in experiment by the interlayer distance – one can distinguish between two regimes. At large separation, the interlayer interaction can be treated perturbatively. When the two layers are brought closer together, the interlayer interaction can become strong enough to lead to the formation of states with *interlayer* correlations. In this paper, we consider the regime of weak interlayer interaction.

The picture of momentum transfer between the two layers leads to the following basic predictions. Clearly, the momentum transfer is independent

---

\* Also at A.F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia.

\*\* Also at Petersburg Nuclear Physics Institute, 188350 St. Petersburg, Russia.

of the nature (electrons or holes) of the carriers in the two layers. Thus, one expects that the sign of the drag resistance depends on the relative sign of the carriers in the two layers. Conventionally, one refers to the resulting drag between two layers with equal (opposite) carriers as positive (negative). A consequence of this is that drag vanishes for particle-hole symmetric systems. In addition, drag is expected to increase monotonously with temperature due to the increased phase space for interlayer scattering. These basic expectations were well supported by early theoretical considerations of Zheng and MacDonald [3].

## 1 Coulomb Drag in High Landau Levels

Remarkably, a series of experiments [4,5,6] on drag in quantum Hall systems shows that these simple expectations are violated in the regime of high Landau levels (large filling factor). First experiments [4,5] showed that drag can change sign as function of magnetic field or filling factor difference, when the two layers are at different densities. This was surprising because the carriers in both layers are electrons. A more recent experiment [6] studied the temperature dependence and found that while exhibiting the expected increase with temperature at high temperatures, an additional peak develops at low temperatures which can be either positive or negative, depending on the filling factor difference between the two layers. This non-monotonous temperature dependence contrasts with the phase-space argument given above.

Drawing an analogy between partially filled Landau levels and partially filled Bloch bands, it is natural to suggest [4] that a less-than-half-filled Landau level behaves as an electron-like band while a more than half-filled Landau level behaves as a hole-like band. This suggests that there are sign changes in the drag resistivity as function of the filling factor difference between the two layers. While there is an element of truth in this picture, it turns out [7] that this analogy can be rather misleading. One reason is that it neglects the contributions to the drag resistivity due to the Hall conductivity.

We find [8] from a careful evaluation of the diagrammatic expressions for the drag conductivity that the experiments can be understood in terms of the interplay of two contributions which are associated with two sources of particle-hole symmetry violation. One source is the curvature of the (quadratic) zero-magnetic-field dispersion of the electrons. This contribution is closely related to the conventional picture for drag described above. Thus, this contribution which dominates at high temperatures, indeed increases monotonously with temperature and has a fixed sign (positive for two electron layers). A second source is related to the oscillatory Landau-level density of states. This anomalous contribution dominates at low temperatures where it exhibits a peak. Its sign depends on the filling factor difference between the two layers. In the ballistic regime  $R_c \gg d$  ( $R_c$  denotes the cyclotron radius and  $d$  the

spacing between the two layers), the sign turns out to be in agreement with experiment.

## 2 Previous Theoretical Work

We briefly discuss previous theoretical work. Early theoretical work [9] on Coulomb drag in a magnetic field in the limit of high Landau levels showed that the magnetic field strongly enhances the Coulomb drag. More recent work [7] showed that Landau-level quantization can lead to sign changes in drag. However, the results obtained for the diffusive regime indicated that unlike the experimental observation, negative drag should be observed for equal filling factors in the two layers. None of these works can explain the observed non-monotonic temperature dependence.

Ref. [7] started from an alternative picture for the drag *conductivity* which can be roughly thought of as follows. The fluctuations of the electron density of the active layer lead to a fluctuating electric potential which is seen by the passive layer. Due to the applied voltage in the active layer, these electric potential fluctuations are not fully isotropic in the wavevector  $\mathbf{q}$  and thus induce a *dc* current in the passive layer in nonlinear response. This argument shows that drag is related to the "nonlinear susceptibility"  $\Gamma(\mathbf{q}, \omega)$  which, *loosely speaking*, gives the *dc* current response to an electric potential  $\phi(\mathbf{q}, \omega)$  in quadratic order,  $j_{dc} = \Gamma(\mathbf{q}, \omega)|\phi(\mathbf{q}, \omega)|^2$ . (A more accurate, diagrammatic definition is given in the next section.) The diagrammatic approach (see below) shows that the anisotropic part of the electric potential fluctuations of the active layer is also related to  $\Gamma(\mathbf{q}, \omega)$  of the active layer. As a result, one finds for the *dc* drag conductivity the expression [10,11]

$$\sigma_{ij}^{(D)} = \frac{e^2}{16\pi TS} \sum_{\mathbf{q}} \int_{-\infty}^{\infty} \frac{d\omega}{\sinh^2(\omega/2T)} \Gamma_i^{(1)}(\mathbf{q}, \omega) \Gamma_j^{(2)}(\mathbf{q}, \omega) |U(\mathbf{q}, \omega)|^2. \quad (1)$$

where  $S$  denotes the area of the sample,  $T$  the temperature, and  $U(\mathbf{q}, \omega)$  the screened interlayer interaction.

In the limit in which the current response of the passive layer to the electric potential fluctuation generated by the active layer is local, one can argue [12,7] that the "nonlinear susceptibility" becomes

$$\Gamma(\mathbf{q}, \omega) = 2 \frac{d\sigma}{d(en)} \cdot \mathbf{q} \text{Im}\Pi(\mathbf{q}, \omega), \quad (2)$$

where  $\Pi(\mathbf{q}, \omega)$  is the polarization operator of the layer,  $\sigma$  the conductivity tensor, and  $n$  denotes the electron density. Inserting this into the expression for the drag conductivity which is related to the experimentally measured drag resistivity by  $\rho_D = \rho^{(1)}\sigma^{(D)}\rho^{(2)}$  (here  $\rho^{(i)}$  denotes the resistivity of the  $i$ th layer), one finds [7] that the drag resistivity, takes the form

$$\rho^{(D)} \sim \rho^{(1)} \frac{d\sigma^{(1)}}{dn} \frac{d\sigma^{(2)}}{dn} \rho^{(2)}. \quad (3)$$

Here, the prefactor is positive. From this expression, we can read off the following results:

- (1) Replacing the conductivity and resistivity tensors by their Drude form, assuming that the scattering time is independent of density, and using that in this case  $d\sigma/dn = \pm\sigma/n$  (the plus (minus) sign is valid for electrons (holes), respectively), we find that the drag resistivity is diagonal and its sign is in agreement with the simple arguments in the introduction to this paper.
- (2) Even for the Drude conductivities, a density-dependent scattering time results in a finite Hall drag, i.e., an off-diagonal contribution to the drag resistivity.
- (3) In the presence of a strong magnetic field, the longitudinal conductivity oscillates as a function of  $n$  due to the Shubnikov-de-Haas oscillations. This leads to sign changes in the drag resistivity as function of the density difference between the layers.
- (4) In strong magnetic fields, we have  $\rho_{xx} \ll \rho_{xy}$  and in addition, the self-consistent Born approximation (SCBA) valid in the limit of high Landau levels yields that  $d\sigma_{xx}/dn \gg d\sigma_{xy}/dn$ , so that to leading order

$$\rho_{xx}^{(D)} \sim \rho_{xy}^{(1)} \frac{d\sigma_{yy}^{(1)}}{dn} \frac{d\sigma_{yy}^{(2)}}{dn} \rho_{yx}^{(2)}. \quad (4)$$

Remarkably, since  $\rho_{xy} = -\rho_{yx}$ , this expression predicts that the drag resistivity is *negative* for *equal* densities. Thus, while the expression does exhibit sign oscillations as a function of filling factor differences between the two layers, the overall sign is in stark contrast to the experimentally observed positive sign for equal filling factors.

Thus, we have to go beyond the local, diffusive limit discussed in this section in order to understand the experiments. The corresponding diagrammatic calculation for the ballistic regime has recently been performed by us [8] and is sketched in the next section. We will see that this calculation resolves the discrepancy with experiment found here.

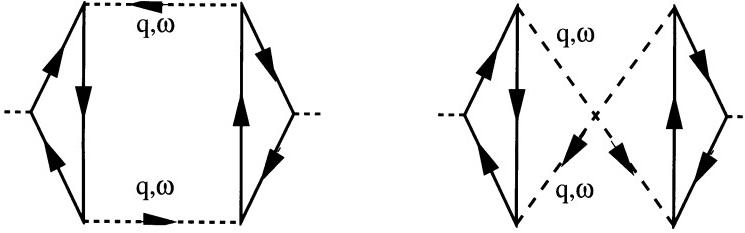
### 3 Diagrammatic Theory

Our considerations [8] are based on the Kubo approach to Coulomb drag [10,11] which expresses the drag conductivity  $\sigma_{ij}^{(D)}(\mathbf{Q}, \Omega)$  in terms of a current-current correlation function,

$$\sigma_{ij}^{(D)}(\mathbf{Q}, \Omega) = \frac{1}{\Omega S} \int_0^\infty dt e^{i\Omega t} \left\langle [j_i^{(1)\dagger}(\mathbf{Q}, t), j_j^{(2)}(\mathbf{Q}, 0)] \right\rangle. \quad (5)$$

where  $i, j$  label the components of the drag conductivity tensor,  $\mathbf{Q}, \Omega$  denote the wave vector and frequency of the applied field,  $S$  is the area of the sample, and  $j_i^{(l)}$  denotes the  $i$ th component of the current operator in the  $l$ th layer. The *dc* drag conductivity follows by taking the limit

$$\sigma_{ij}^{(D)} = \sigma_{ij}^{(D)}(\mathbf{Q} = 0, \Omega \rightarrow 0). \quad (6)$$



**Fig. 1.** Diagrams for the drag conductivity and for the triangle vertex  $\Gamma(\mathbf{q}, \omega)$

Considering the diagrams in Fig. 1, one obtains Eq. (1) for the *dc* drag conductivity [11,10]. The triangle vertex  $\Gamma(\mathbf{q}, \omega)$  takes the explicit form  $\Gamma = \Gamma^{(a)} + \Gamma^{(b)}$  with the two contributions

$$\begin{aligned} \Gamma^{(a)}(\mathbf{q}, \omega) &= \int \frac{d\epsilon}{4\pi i} \tanh \frac{\epsilon + \omega}{2T} \text{tr} \left\{ \mathbf{v} \mathcal{G}^+(\epsilon + \omega) e^{i\mathbf{qr}} \mathcal{G}^+(\epsilon) e^{-i\mathbf{qr}} \mathcal{G}^+(\epsilon + \omega) \right. \\ &\quad \left. - \mathbf{v} \mathcal{G}^-(\epsilon + \omega) e^{i\mathbf{qr}} \mathcal{G}^-(\epsilon) e^{-i\mathbf{qr}} \mathcal{G}^-(\epsilon + \omega) \right\} + (\omega, \mathbf{q} \rightarrow -\omega, -\mathbf{q}) \end{aligned} \quad (7)$$

$$\begin{aligned} \Gamma^{(b)}(\mathbf{q}, \omega) &= \int \frac{d\epsilon}{4\pi i} \left( \tanh \frac{\epsilon + \omega}{2T} - \tanh \frac{\epsilon}{2T} \right) \\ &\quad \times \text{tr} \left\{ \mathbf{v} \mathcal{G}^-(\epsilon + \omega) e^{i\mathbf{qr}} [\mathcal{G}^-(\epsilon) - \mathcal{G}^+(\epsilon)] e^{-i\mathbf{qr}} \mathcal{G}^+(\epsilon + \omega) \right\} \\ &\quad + (\omega, \mathbf{q} \rightarrow -\omega, -\mathbf{q}) \end{aligned} \quad (8)$$

Here,  $\mathcal{G}^\pm(\epsilon)$  denotes the advanced/retarded Green function for a particular impurity configuration.

We perform the average over the white-noise impurity potential, characterized by  $\langle V(\mathbf{r}) \rangle = 0$  and  $\langle V(\mathbf{r})V(\mathbf{r}') \rangle = (1/2\pi\nu_0\tau_0)\delta(\mathbf{r} - \mathbf{r}')$ , within the self-consistent Born approximation (SCBA) [13]. ( $\nu_0$  and  $\tau_0$  denote the density of states and the elastic scattering time at zero magnetic field, respectively.) This approximation which neglects diagrams with crossing impurity lines, can be shown to give the leading contribution when the Fermi energy  $E_F$  is in a high LL with LL index  $N \gg 1$  [14].

### 3.1 The Triangle Vertex

In the ballistic regime for well-separated Landau levels, we can expand in three small parameters simultaneously. The parameter  $\Gamma/\hbar\omega_c \ll 1$  describes the fact that the LLs are well separated. Here  $\Gamma$  denotes the broadening of the LL and  $\omega_c$  is the cyclotron frequency. The limit of high LLs allows us to treat  $1/N \ll 1$  as a small parameter. Finally, the ballistic regime is characterized by the small parameter  $1/qR_c \ll 1$ .

In the leading order in all three parameters, the following simplifications can be made: (1) There is no vertex correction due to disorder of the vector vertex in  $\Gamma$ . This is actually generally true for white-noise disorder. (2) There are also no vertex corrections of the scalar vertices due to the small parameter  $1/qR_c \ll 1$ . (3) Two of the three Green function can be evaluated in the

$N$ th LL in which the Fermi energy resides. One Green function must be evaluated in a neighboring LL  $N \pm 1$  since the velocity operator, entering at the vector vertex, has nonzero matrix elements between neighboring LLs only. (4) The matrix elements associated with the vertices can be evaluated in the semiclassical approximation due to  $N \gg 1$ .

Remarkably, at this order we find that there is an *exact cancellation* between the contributions  $\Gamma^{(a)}$  and  $\Gamma^{(b)}$ . This cancellation requires us to go beyond the leading order to find the relevant expression for the triangle vertex. Since all three small parameters are of comparable magnitude, we need to look for the leading correction in each of them separately.

The leading correction in the ballistic parameter  $1/qR_c$  requires to retain the vertex corrections to the scalar vertices. Within SCBA, the vertex corrections arise from impurity ladders dressing both scalar vertices. We find

$$\begin{aligned} \Gamma^{(1/qR_c)}(\mathbf{q}, \omega) = & -\hat{\mathbf{q}} \frac{128\omega R_c}{\pi^2 \ell^2} \\ & \times \frac{(E - E_N)[\Gamma^2 - (E - E_N)^2]^{3/2}}{\Gamma^6} J_1(qR_c) J_0^3(qR_c). \end{aligned} \quad (9)$$

Here, we introduced a superscript on  $\Gamma(\mathbf{q}, \omega)$  in order to distinguish this contribution from other contributions computed below. Note that this is a purely longitudinal contribution to  $\Gamma$  (i.e., parallel to  $\mathbf{q}$ ) which changes sign at the center of the LL.

Corrections of order  $\Gamma/\hbar\omega_c$  can arise from two sources: (1) The Green functions adjacent to the current vertex are evaluated in Landau levels different from  $N$ . The Green function between the scalar vertices must still be evaluated in the  $N$ th Landau level because  $G_n^+ - G_n^- \sim \Gamma/(\hbar\omega_c)^2$  for  $n \neq N$ . (2) The diagrams giving the leading contribution can be evaluated more accurately, keeping corrections in  $\Gamma/\hbar\omega_c$ . We may now neglect vertex corrections at the scalar vertices because we consider the leading order in  $qR_c \gg 1$ .

It turns out that the first contribution vanishes for both  $\Gamma^{(a)}$  and  $\Gamma^{(b)}$ . Considering the second contribution, it turns out that it still gives a vanishing contribution to the longitudinal triangle vertex, due to the cancellation between  $\Gamma^{(a)}$  and  $\Gamma^{(b)}$  described above. However, the transverse contribution to  $\Gamma^{(b)}$  no longer vanishes when considering corrections in  $\Gamma/\hbar\omega_c$ . In this way, we obtain the contribution

$$\begin{aligned} \Gamma^{(\Gamma/\hbar\omega_c)}(\mathbf{q}, \omega) = & -\hat{\mathbf{q}} \times \hat{\mathbf{z}} \frac{8\omega R_c}{\pi^2 \ell^2} \\ & \times \frac{(E - E_N)[\Gamma^2 - (E - E_N)^2]}{\omega_c \Gamma^4} J_0(qR_c) J_1(qR_c) \end{aligned} \quad (10)$$

to the triangle vertex. This is a transverse contribution to  $\Gamma$  which also changes sign in the center of the LL.

Finally, we consider the contribution to  $\Gamma$  due to terms of order  $q/k_F$  relative to the leading order. Such terms arise from a more accurate treatment of the matrix elements involved in the scalar vertices, keeping corrections of

order  $1/N$  in the arguments of oscillatory terms. These corrections arise only for the contribution  $\Gamma^{(b)}$  and one finds

$$\Gamma^{(q/k_F)} = -\hat{\mathbf{z}} \times \mathbf{q} \frac{4\omega}{\pi^2} \frac{\Gamma^2 - (E - E_N)^2}{\Gamma^4} J_0^2(qR_c). \quad (11)$$

This expression can also be rewritten as

$$\Gamma^{(q/k_F)} = -\hat{\mathbf{z}} \times \mathbf{q} \frac{2\sigma_{xy}}{n} \text{Im}\Pi(\mathbf{q}, \omega) \quad (12)$$

with the polarization operator  $\Pi(\mathbf{q}, \omega)$  for the ballistic regime [cf. Eq. (15) below] and the Hall conductivity  $\sigma_{xy} = en_2/B - (e^2/\pi^2\hbar)N(\Gamma/\hbar\omega_c)[1 - (E - E_N)^2/\Gamma^2]^{3/2}$  in SCBA. This is a high-magnetic field analog of the conventional contribution to  $\Gamma$  [3] and does not change sign in the center of the LL.

### 3.2 Screened Interlayer Interaction

In this section, we briefly summarize the results for the screened interlayer interaction. For  $q$  small compared to the Thomas-Fermi screening wave vectors  $\kappa_{l,0} = 4\pi e^2 \nu_{l,0}$  ( $l = 1, 2$  labels the layer and  $\nu_{l,0}$  denotes the zero-field density of states per spin of layer  $l$ ), the screened interlayer interaction can be approximated as [11]

$$U(\mathbf{q}, \omega) \simeq \frac{\pi e^2 q}{\kappa_{1,0} \kappa_{2,0} \sinh(qd)} \frac{2\nu_{1,0}}{\Pi_1(\mathbf{q}, \omega)} \frac{2\nu_{2,0}}{\Pi_2(\mathbf{q}, \omega)}. \quad (13)$$

Here,  $d$  denotes the distance between the layers. The polarization operator of layer  $l$  is denoted by  $\Pi_l(\mathbf{q}, \omega)$ .

In the ballistic regime  $qR_c \ll 1$ , for  $\omega \ll \Gamma$ , the real part of the polarization operator takes the form

$$\text{Re}\Pi(\mathbf{q}, \omega) = 2\nu_0 + 2\nu(E)[J_0(qR_c)]^2. \quad (14)$$

Here, the second term arises from contributions of the  $N$ th LL, while the first term [15] arises from Landau levels  $n \neq N$ . The imaginary part of the polarization operator takes the form [16]

$$\text{Im}\Pi(\mathbf{q}, \omega) = 2\nu_0 \frac{2\omega\omega_c}{\pi\Gamma^2} \left[ 1 - \frac{(E - E_n)^2}{\Gamma^2} \right] [J_0(qR_c)]^2 \quad (15)$$

Note that  $\text{Im}\Pi \ll \text{Re}\Pi$ .

### 3.3 Results for the Drag Resistivity

In this section, we summarize our results for the drag resistivity based on the formalism sketched above [8]. In strong magnetic fields, the Hall conductivity

$\rho_{xy}$  dominates over  $\rho_{xx}$ . Therefore, using Eq. (1), the drag resistivity  $\rho_{xx}^{(D)} \simeq \rho_{xy}^{(1)} \sigma_{yy}^D \rho_{yx}^{(2)}$  can be expressed as

$$\begin{aligned} \rho_{xx}^D = & -\frac{B}{en_1} \frac{B}{en_2} \frac{1}{8\pi} \int_{-\infty}^{\infty} \frac{d\omega}{\sinh^2(\omega/2T)} \\ & \times \int \frac{d^2\mathbf{q}}{(2\pi)^2} \Gamma_y^{(1)}(\mathbf{q}, \omega, B) \Gamma_y^{(2)}(\mathbf{q}, \omega, -B) |U(\mathbf{q}, \omega)|^2. \end{aligned} \quad (16)$$

A careful analysis shows that at low temperatures,  $T \ll \Gamma$ , the contribution of  $\Gamma^{(\Gamma/\hbar\omega_c)}$  to  $\rho^{(D)}$  dominates. Performing the remaining integrals yields the final result

$$\begin{aligned} \rho_{xx}^D = & \frac{8}{3\pi^2 e^2} \frac{1}{(k_F d)^2} \frac{1}{(\kappa R_c)^2} \left( \frac{T}{\Gamma} \right)^2 \ln \left( \frac{R_c \Gamma}{d\omega_c} \right) \\ & \times \left( \frac{E - E_N}{\Gamma} \right)^2 \left[ 1 - \frac{(E - E_N)^2}{\Gamma^2} \right]^2. \end{aligned} \quad (17)$$

for the drag resistivity of identical layers. Thus at low temperature  $T \ll \Gamma$ , the drag resistivity scales with magnetic field and temperature as  $\rho_{xx}^D \propto T^2 B \ln B$ , which compares nicely to the empirical scaling reported in [6],  $\rho_{xx}^D \propto (n/B)^{-2.7} f(T/B)$ , with  $f(x) \sim x^2$ .

For different filling factors (or densities) in the two layers, the cyclotron radii of the two layers are no longer the same. If the cyclotron radii  $R_c$  of the two layers are slightly different, such that  $\delta R_c/d \ll 1$ , the above calculation fully applies, and the result is obtained by the replacement

$$\begin{aligned} & \left( \frac{E - E_N}{\Gamma^2} \right)^2 \left[ 1 - \frac{(E - E_N)^2}{\Gamma^2} \right]^2 \\ & \rightarrow \left( \frac{E - E_N}{\Gamma} \left[ 1 - \frac{(E - E_N)^2}{\Gamma^2} \right] \right)_1 \left( \dots \right)_2 \end{aligned} \quad (18)$$

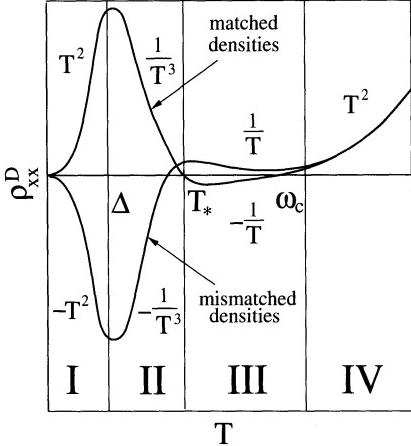
in Eq. (17). This yields an *oscillatory sign* of the drag resistivity. Importantly, for equal layer densities the drag resistivity is *positive*, in agreement with experiment. This result is different from the result for the diffusive regime discussed above [7]. The origin of the difference between the diffusive and the ballistic regime is that in the former (latter), the leading term originates from a longitudinal (transverse) contribution to  $\mathbf{\Gamma}$ .

If  $d \ll \delta R_c \ll R_c \Gamma / \omega_c$ , the calculation still applies if we replace in (17)

$$\ln \left( \frac{R_c \Gamma}{d \omega_c} \right) \rightarrow \ln \left( \frac{R_c \Gamma}{\delta R_c \omega_c} \right). \quad (19)$$

For larger temperatures  $T \gg \Gamma$ , we find a number of different temperature regimes which are summarized in Fig. 2. For  $\Gamma \ll T \ll T_*$  (regime II in Fig. 2), we find that the drag resistivity scales as

$$\rho_{xx}^D \propto T^{-3} B^{7/2} \ln B \quad (20)$$



**Fig. 2.** Schematic temperature dependence of drag for matched and mismatched densities. In the latter case, the mismatch is chosen such that drag is negative at low temperatures

as function of  $B$  and  $T$ . This contribution arises from  $\Gamma^{(\Gamma/\hbar\omega_c)}$ . The upper limit of the temperature interval is given by  $T_* \equiv \Gamma(\Gamma/\hbar\omega_c)k_F d$ . For even higher temperatures, the conventional contribution  $\Gamma^{(q/k_F)}$  dominates and the drag resistivity becomes positive. Specifically, we find that the drag resistivity scales as  $\rho_{xx}^D \propto T^{-1}B^{7/2}$  for  $T_* \ll T \ll \omega_c$  (regime III in Fig. 2) and

$$\rho_{xx}^D \sim \frac{1}{e^2} \frac{1}{(k_F a)^2} \frac{1}{(\kappa R_c)^2} \left( \frac{T}{E_F} \right)^2 \left( \frac{\omega_c}{T} \right) \propto T^2 B^{1/2}. \quad (21)$$

for even higher temperatures  $T \gg \omega_c$  (region IV in Fig. 2).

## 4 Summary and Outlook

Experimental studies of drag in high Landau levels have given a number of surprises, including negative drag and a non-monotonous temperature dependence. Earlier theoretical work had led to some understanding of the possibility of sign changes of drag in this regime, but the detailed predictions for the sign did not agree with experiment. In this contribution, we sketch our results for Coulomb drag in high Landau levels, obtained in the framework of a diagrammatic calculation. Our calculation treats the interlayer interaction perturbatively and accounts for disorder in the self-consistent Born approximation which is expected to become exact in the limit of high Landau levels. Focusing on the ballistic regime for well-separated Landau levels, we find that there are two types of contributions to the triangle vertex entering the expression for the drag resistivity which are associated with different sources for a particle-hole asymmetry. At high temperatures, the contribution arising from the particle-hole asymmetry due to the zero-field electron dispersion dominates. This contribution is positive and increases with temperature. At lower temperatures, an anomalous contribution dominates which arises from

the particle-hole asymmetry associated with the Landau level spectrum. This contribution changes sign in a way which is consistent with experimental observations. We close by mentioning that it would be interesting to apply the approach outlined here to the fractional quantum Hall regime close to half filling of the lowest Landau level.

### Acknowledgements

We thank J.G.S. Lok, K. von Klitzing, A. Khaetskii, and Ady Stern for useful discussions. One of us thanks the Weizmann Institute for hospitality during the preparation of this manuscript (supported by LSF grant HPRI-CT-2001-00114 and the Einstein Center). This work has been supported by the DFG-Schwerpunkt “Quanten-Hall-Systeme” (FvO, IVG, ADM), SFB 290 (FvO), as well as the “Junge Akademie” (FvO).

### References

1. T.J. Gramila, et al., Phys. Rev. Lett. **66**, 1216 (1991).
2. U. Sivan, P.M. Solomon, and H. Shtrikman, Phys. Rev. Lett. **68**, 1196 (1992).
3. L.Zheng and A.H. MacDonald, Phys. Rev. B **48**, 8203 (1993).
4. X.G. Feng, et al., Phys. Rev. Lett. **81**, 3219 (1998).
5. J.G.S. Lok, et al., Phys. Rev. B **63**, 041305 (2001).
6. K. Muraki, et al., cond-mat/0311151 (2003).
7. F. von Oppen, S.H. Simon, A. Stern, Phys. Rev. Lett. **87**, 106803 (2001).
8. I.V. Gornyi, A.D. Mirlin, and F. von Oppen, in preparation (2004).
9. M.C. Bønsager, et al., Phys. Rev. B **56**, 10314 (1997).
10. K. Flensberg, et al., Phys. Rev. B **52**, 14761 (1995).
11. A. Kamenev and Y. Oreg, Phys. Rev. B **52**, 7516 (1995).
12. B. N. Narozhny, I. L. Aleiner, Ady Stern, Phys. Rev. Lett. **86**, 3610 (2001).
13. T. Ando and Y. Uemura, J. Phys. Soc. Jpn. **36**, 959 (1974); T. Ando, *ibid* **37**, 1233 (1974).
14. K.A. Benedict and J.C. Chalker, J. of Physics C **19**, 3587 (1986).
15. I.L. Aleiner and L. Glazman, Phys. Rev. B **52**, 11296 (1995).
16. A. Khaetskii and Yu. Nazarov, Phys. Rev. B **59**, 7551 (1999).

## Part IV

### **Quantum Dots, Wires and Rings**

# Applications of Adiabatic Passage in Solid-State Devices

Jens Siewert<sup>1</sup> and Tobias Brandes<sup>2</sup>

<sup>1</sup> Institut für Theoretische Physik, Universität Regensburg  
93040 Regensburg, Germany

<sup>2</sup> Department of Physics, UMIST  
Manchester M60 1QD, UK

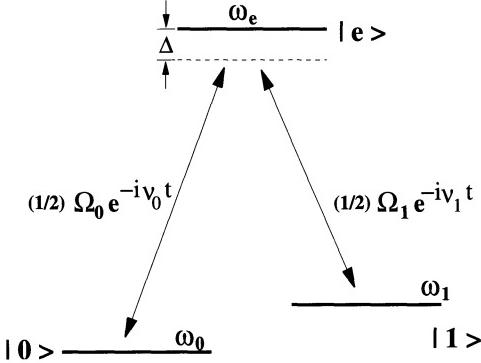
**Abstract.** Coherent population transfer by adiabatic passage is a well-known method in quantum optics. This remarkable technique which is based on simple ideas has remained largely unknown to solid-state physicists. Here we provide an introduction to the basic principles of this method and discuss also some applications in solid-state systems.

## 1 Introduction

Traditionally solid-state physics is focused on systems with quasi-continuous energy spectra for the various elementary excitations present in macroscopic solids. During the past decade, however, the successful development of nanotechnology has made it possible to study devices below the micrometer scale, to manipulate single charges and fluxes, or to build small electrical and mechanical resonators. The typical energy scales for such devices are, say, on the order of 0.1 ... 1 Kelvin and therefore much larger than the smallest temperatures that can be reached in an experiment. Consequently, the discrete nature of the energy spectrum needs to be taken into account in order to understand the behavior of these devices.

Another interesting aspect of solid-state systems at the nanoscale is that they may exhibit quantum coherence phenomena. These new effects have attracted significant attention in the recent past. Controlling quantum coherence in solid-state nanodevices is one of the major objectives in present-day research. Achievement of this goal would make it possible to study quantum dynamics on a macroscopic scale, to provide better insight into the mechanisms of decoherence, and possibly to realize quantum information processing in practice.

A consequence of this evolution is that methods used, e.g., in quantum optics become directly relevant also in solid-state physics. A particularly interesting technique is the so-called stimulated Raman adiabatic passage (STIRAP) that has been developed by Bergmann and co-workers [1,2]. This method can be used to change the quantum state of a system by controlling certain coupling parameters. The mathematics underlying this technique is



**Fig. 1.** Three-level atom with two ground states  $|0\rangle$ ,  $|1\rangle$  coupled to an excited state  $|e\rangle$  via laser fields  $(1/2)\Omega_0 \exp(-i\nu_0 t)$ ,  $(1/2)\Omega_1 \exp(-i\nu_1 t)$ . Note that there is no direct coupling between the states  $|0\rangle$  and  $|1\rangle$

rather simple and general such that it may apply to physically completely different situations.

Until now, adiabatic passage has had only very few applications in solid-state devices e.g. [3,4,5,6]. However, one may hope that this method, due to its elegance and simplicity, will attract more interest and find new applications to solid-state systems in the near future. In this contribution we provide a brief introduction to the basic principles of adiabatic passage. In Section 2 we will explain population transfer in three-level atoms, as it is known from quantum optics. In Section 3, the method is ‘translated’ to a simple solid-state device, a superconducting Cooper-pair box with two islands. We will show that adiabatic charge transfer between the islands is possible in close analogy to the three-level atom. Finally, in Section 4 we will briefly discuss two advanced applications that are based on the technique of adiabatic passage. In our opinion, these applications may serve to illustrate the enormous potential of the method in the field of condensed-matter physics.

## 2 Adiabatic Passage in Three-Level Atoms

In this section, we briefly discuss the basics of adiabatic passage. A more complete discussion of the underlying physics can be found, e.g., in the review articles Refs. [1,2] and in the textbook Ref. [7].

Consider an atom with a  $\Lambda$ -type three-level configuration as shown in Fig. 1. The long-lived ground states  $|0\rangle$  and  $|1\rangle$  (energies  $\omega_0$ ,  $\omega_1$ ) are coupled to an excited state  $|e\rangle$  (energy  $\omega_e$ ) via (classical) laser fields with Rabi frequencies  $\Omega_0$ ,  $\Omega_1$ . The laser frequencies are assumed to have the same detuning  $\Delta$  with respect to the atomic transitions

$$\Delta = (\omega_e - \omega_0) - \nu_0 = (\omega_e - \omega_1) - \nu_1 . \quad (1)$$

If the state  $|\psi(t)\rangle$  of the atom is written in the form

$$|\psi(t)\rangle = c_e(t)e^{-i(\omega_e - \Delta)t}|e\rangle + c_0(t)e^{-i\omega_0 t}|0\rangle + c_1(t)e^{-i\omega_1 t}|1\rangle , \quad (2)$$

the equation of motion for the system (atom + laser fields) in the rotating frame reads

$$\begin{aligned} i\dot{c}_e &= (1/2)(2\Delta c_e + \Omega_0 c_0 + \Omega_1 c_1) \\ i\dot{c}_0 &= (1/2) \Omega_0 c_e \\ i\dot{c}_1 &= (1/2) \Omega_1 c_e \end{aligned} \quad (3)$$

where we have used the rotating-wave approximation and neglected finite lifetimes of the atomic levels. The corresponding Hamiltonian written in the basis  $\{|e\rangle, |0\rangle, |1\rangle\}$  is

$$H = \frac{1}{2} \begin{bmatrix} 2\Delta & \Omega_0 & \Omega_1 \\ \Omega_0 & 0 & 0 \\ \Omega_1 & 0 & 0 \end{bmatrix}. \quad (4)$$

The eigenenergies of this Hamiltonian are easily found to be

$$E^0 = 0 \quad \text{and} \quad E^\pm = \frac{1}{2} \left( \Delta \pm \sqrt{\Delta^2 + \Omega_0^2 + \Omega_1^2} \right). \quad (5)$$

Note that one of the two zero diagonal elements of the Hamiltonian (4) survives *independently* of the couplings  $\Omega_j$  and the offset  $\Delta$ . The corresponding eigenstate (which is *not* the ground state) reads

$$|\phi^0\rangle = \frac{1}{\sqrt{\Omega_0^2 + \Omega_1^2}} (\Omega_1 |0\rangle - \Omega_0 |1\rangle). \quad (6)$$

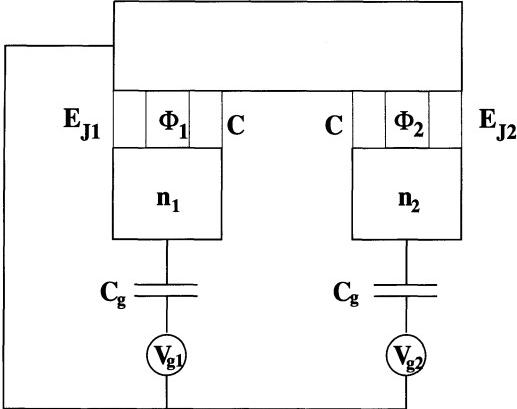
The remarkable property of this state is that it does not contain an admixture of the basis vector  $|e\rangle$ . Therefore it cannot decay by spontaneous emission from  $|e\rangle$  and is usually called a ‘dark state’. The direction of  $|\phi^0\rangle$  in the subspace  $\{|0\rangle, |1\rangle\}$  is given by the (possibly complex) coupling parameters  $\Omega_j$ .

Now consider the case of weakly time-dependent coupling parameters. If the inverse characteristic scale of this time dependence is smaller than the level spacing  $\Omega_{\text{eff}}$  of the Hamiltonian (4)

$$\left| \frac{\dot{\Omega}_j}{\Omega_j} \right| \ll \Omega_{\text{eff}} \sim |\Omega_j| \quad (7)$$

the adiabatic theorem guarantees that a given state follows the time evolution of the coupling coefficients. Consequently, the dark state can be rotated in the subspace  $\{|0\rangle, |1\rangle\}$  by slowly changing the values of the couplings. Again, this is remarkable as the *quantum* state can be controlled by varying *classical* system parameters.

Population transfer from the state  $|0\rangle$  to the state  $|1\rangle$  can be achieved by applying the so-called counterintuitive scheme. The system is prepared in the state  $|0\rangle$  at vanishing couplings. Switching on the laser 1 to a finite  $\Omega_1$  does not affect the state of the system. Now, also the laser 0 is slowly switched on (finite  $\Omega_0$ ) while  $\Omega_1$  is turned off. As can be immediately read off Eq. (6) the dark state  $|\phi^0\rangle$  starts to rotate towards the state  $|1\rangle$ . At the end of the switching procedure we have  $\Omega_1 = 0$ , finite  $\Omega_0$ , and the final state equals  $|1\rangle$ .



**Fig. 2.** Circuit for a Cooper-pair box with two islands. For simplicity we choose equal gate capacitances  $C_g$  and also equal junction capacitances  $C$ . The system is operated in the charging regime  $E_C \gg E_{Jj}$ . The  $\Lambda$  scheme can be recognized by looking at the position of the excess Cooper pair (for a detailed explanation, see text):  $|0\rangle$  = pair in the left island,  $|1\rangle$  = pair in the right island,  $|e\rangle$  = pair in the lead

### 3 Analogy for a Cooper-Pair Box with Two Islands

Our aim is to show that Hamiltonians similar to Eq. (4) can be ‘tailored’ in solid-state devices. As an example we consider a superconducting Cooper-pair box with two islands. The (single-island) Cooper-pair box is a well-known system as it is a promising candidate for the practical realization of a solid-state qubit (see, e.g., [8] and references therein).

The circuit of a two-island Cooper-pair box is shown in Fig. 2. The superconducting islands are coupled via tunable Josephson junctions to a (large) superconducting lead. The SQUID-loop layout of the junctions makes it possible to control the coupling energies  $E_{J1}, E_{J2}$  by means of the external fluxes  $\Phi_1, \Phi_2$ . The electrostatic potentials of the islands can be changed through the gate voltage sources  $V_{gj}$  ( $j = 1, 2$ ) which induce offset charges  $n_{xj} = C_{gj}V_{gj}/2e$  on the islands (in units of Cooper-pair charges).

Let us calculate the energy of the circuit in Fig. 2 according to classical electrostatics. We assume  $C_{g1} = C_{g2} = C_g$  and  $C_1 = C_2 = C$  and abbreviate  $C_\Sigma \equiv 2C + 2C_g$ . Then the electrostatic energy can be written in terms of the number of island charges  $n_j$  (more precisely: excess charges) as

$$\begin{aligned} \mathcal{E}(n_1, n_2) &= E_C(n_1 - n_{x1})^2 + E_C(n_2 - n_{x2})^2 , \\ &\longrightarrow E_C(N - N_x)^2 + E_C[(n_1 - n_2) - (n_{x1} - n_{x2})]^2 + \text{const.} \end{aligned} \quad (8)$$

where  $E_C = (2e)^2/(2C_\Sigma)$  denotes the charging energy scale. In the second line we have written  $N \equiv n_1 + n_2$  for the total charge number on both islands (and correspondingly  $N_x \equiv n_{x1} + n_{x2}$ ). Observe that the number of island charges  $n_j$  are *discrete* variables.

Assume we choose  $N_x \simeq 1/2$ . Then, the charge states with the lowest electrostatic energy correspond to a total island charge of  $N = 0$  and  $N = 1$ . If we further put  $n_{x1} = n_{x2}$  the charge states with  $n_{x1} = 1$ ,  $n_{x2} = 0$  (and, correspondingly,  $n_{x1} = 0$ ,  $n_{x2} = 1$ ) have the same energy.

In summary, we have obtained that, given the offset charges mentioned above, the three charge states with the lowest electrostatic energy are  $|0\rangle$  with “one charge on island 1, no charge on island 2”,  $|1\rangle$  with “no charge on island 1, one charge on island 2”, and  $|e\rangle$  with zero excess charge on both islands 1 and 2. All other charge states have energies that are on the order of  $E_C$  higher.

Now we include also Josephson tunneling in our discussion. The device is to be operated in the charging regime, that is

$$E_{Jj} \ll E_C . \quad (9)$$

If we are interested in the low-energy dynamics it is sufficient to consider only the lowest-lying charge states and the relevant Josephson couplings. This is in complete analogy with the reasoning for the one-island Cooper-pair box in Ref. [8]. The resulting Hamiltonian that describes the quantum dynamics of the circuit in Fig. 2 is

$$\begin{aligned} H = & \mathcal{E}(0,0)|e\rangle\langle e| + \mathcal{E}(1,0)|0\rangle\langle 0| + \mathcal{E}(0,1)|1\rangle\langle 1| - \\ & -(E_{J1}/2)(|e\rangle\langle 0| + |0\rangle\langle e|) - (E_{J2}/2)(|e\rangle\langle 1| + |1\rangle\langle e|) , \end{aligned} \quad (10)$$

where the first line describes the charging part discussed above, and the second line is the tunneling Hamiltonian in the three-dimensional subspace. As it is possible to choose  $\mathcal{E}(1,0) = \mathcal{E}(0,1)$ , we see that the Hamiltonian Eq. (10) has exactly the same structure as the Hamiltonian Eq. (4) of the three-level atom in the rotating frame.

From the mathematical equivalence of the Hamilton operators in Eqs. (4) and (10) we conclude that an adiabatic population transfer as described in Section 2 is possible also in the two-island Cooper-pair box. Here, switching the coupling parameters means to vary the Josephson tunnel couplings by tuning the local fluxes  $\Phi_1$ ,  $\Phi_2$ . Physically, the population transfer corresponds to moving the excess Cooper pair from island 1 to island 2.

## 4 Solid-State Applications of Adiabatic Passage

In this section, we will illustrate further applications of adiabatic population transfer in solid-state devices. It is evident from these examples that the method provides unexpected solutions to interesting problems, therefore one might hope that it finds a wider range of condensed-matter applications in the future.

#### 4.1 Non-abelian Holonomies by Sequences of Adiabatic Population Transfers

It is certainly an interesting problem to demonstrate the existence of geometric phases and to measure them quantitatively [9]. Numerous manifestations of geometric phases in physics are so-called Berry phases [10]. This phase occurs when a *non-degenerate* quantum state carries out a cyclic evolution due to cyclic adiabatic parameter changes of the Hamiltonian.

The generalization of the Berry phase to *degenerate* states is the non-Abelian holonomy [11]. Consider a quantum system which depends on an  $n$ -tuple of parameters  $\{\lambda_1, \dots, \lambda_n\}$  (external fields, etc.), the *control manifold*. Moreover, let the system have a degenerate subspace which remains degenerate for any parameter point of the control manifold (this is a rather non-trivial assumption). We prepare the system in a state that is an element of the degenerate subspace and perform a cyclic adiabatic evolution of the Hamiltonian along a closed contour in the control manifold.

While a non-degenerate state returns to the initial state (times a phase factor) at the end of the cyclic evolution (a consequence of the adiabatic theorem), a state in the degenerate subspace will, in general, experience a rotation  $U$  within the degenerate subspace. This rotation is called a non-Abelian holonomy. Mathematically  $U$  is given by a path-ordered integral along the contour  $C$

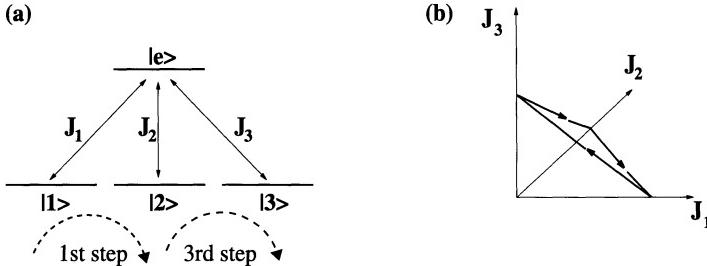
$$U = \mathcal{P} \exp \oint_C \chi \, d\lambda \quad (11)$$

where  $\chi$  is the (matrix-valued) Wilczek-Zee connection [11]. In general this expression is rather difficult to evaluate. Therefore, it is an interesting question whether one can find contours in the control manifold for which one can immediately see the corresponding holonomy  $U$ . This is of relevance also for an experimental demonstration of non-Abelian holonomies.

Adiabatic passage provides an answer to this question that has been found in the context of holonomic quantum computation [12,13,5]. Consider a four-level scheme with states  $|e\rangle, |1\rangle, |2\rangle, |3\rangle$  (see Fig. 4.1a) where the control manifold is given by the coupling parameters  $\{J_1, J_2, J_3\}$ . This system can be realized, e.g., in a superconducting nanocircuit with three islands in analogy with the circuit in Fig. 2 (see also Ref. [5] where a similar system has been studied).

It is easy to show that this system has a two-dimensional degenerate subspace for any parameter configuration in the control manifold. The system is prepared in the state  $|1\rangle$  and  $J_1 = 0, J_2 = \mathcal{J}, J_3 = 0$ . Now we perform a 3-step sequence of adiabatic passages (cf. also Fig. 4.1b):

- 1st step:* switch off  $J_2$  while  $J_1$  is switched on  
 $\longrightarrow$  state of system changes  $|1\rangle \rightarrow |2\rangle$
- 2nd step:* switch off  $J_1$  while  $J_3$  is switched on  
 $\longrightarrow$  state of system does not change



**Fig. 3.** (a) Four-level scheme for the generation of non-Abelian holonomies. The energies of the levels (i.e., the diagonal elements of the  $4 \times 4$  Hamiltonian analogous to Eq. (4)) are  $E_e = \varepsilon$  and  $E_1 = E_2 = E_3 = 0$ . (b) Contour in the control manifold described in the text. The starting point is  $J_1 = J_3 = 0$ ,  $J_2$  finite

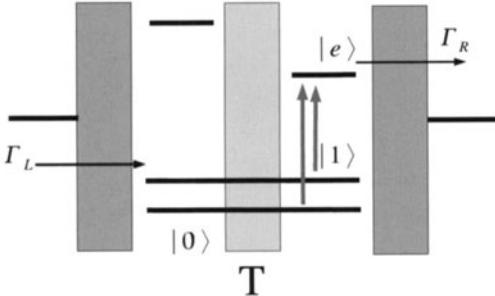
*3rd step:* switch off  $J_3$  while  $J_2$  is switched on  
 $\longrightarrow$  state of system changes  $|2\rangle \rightarrow |3\rangle$

We see that, while a closed contour is described in the control manifold, the state of the system is rotated from state  $|1\rangle$  to state  $|3\rangle$  (possibly times a phase factor). That is, we have found a simple non-Abelian holonomy which can be generated experimentally in a straightforward manner.

## 4.2 Coupled Quantum Dots

Another solid-state application is the realization of dark states and adiabatic passage in coupled quantum dots. The original proposal [3] with two coupled quantum dots in the strong Coulomb blockade regime is actually quite close to three-level systems in atoms, with the additional possibility to test the effect (and its modifications) in electronic transport, cf. Fig. 4. The dark state Eq. (6) appears in the form of a sharp anti-resonance in the *stationary current* through a double dot as a function of the ‘Raman detuning’, i.e. the detuning difference of the two classical laser (or microwave) fields. The half-width of the anti-resonance can then be used to extract valuable information, such as the relaxation and dephasing times of tunnel coupled dot-ground state superpositions, from transport experiments.

Using two time-dependent Stokes and pump pulses, an extension of the counterintuitive STIRAP scheme has subsequently been suggested for this configuration [3], taking into account a finite decoherence rate  $\gamma$  due to, e.g. electron-phonon coupling in the dots. In principle,  $\gamma$  can be obtained from monitoring the time-dependent electronic current through the dots, which however is quite weak for small  $\gamma$ , when the dot is essentially trapped in the dark superposition of  $|0\rangle$  and  $|1\rangle$ . An alternative way is to apply a *second* pair of simultaneous pulses with amplitude ratios that give either zero or full current in the coherent case  $\gamma = 0$ . The deviation from the ‘zero/full’ current situations then gives rise for a current ‘contrast’ from which  $\gamma$  can be extracted.



**Fig. 4.** Dark states and STIRAP through two tunnel ( $T$ ) coupled dots in the Coulomb blockade regime with hybridized states  $|0\rangle$ ,  $|1\rangle$ . Electrons tunnel in from the left, are photo-pumped to the excited state  $|e\rangle$ , and tunnel out to the right

Another adiabatic scheme that completely avoids the use of lasers or microwaves has been introduced in quantum dots which are coupled by slowly varying static tunnel barriers. In fact, the rotating wave approximation in the original (optical) population transfer scheme leads to time-independent (or slowly parametric) Hamiltonians like Eq. (4), where fast terms are already transformed away. This makes it obvious that one can start from time-dependent tunnel couplings right from the beginning. In the simplest realization [4], one considers *three* single level dots  $L$ ,  $C$ ,  $R$  in a line, with two time-dependent couplings  $T_i(t)$  between  $C$  and  $i = L, R$  which are then switched on and off with a time-delay as in the STIRAP scheme. The resulting adiabatic transfer of charge from the left to the center to the right can then essentially be understood in terms of level-crossings of the (instantaneous) three eigenvalues of the energy. If the tunnel coupling remains small but finite in the ‘off’ periods, these level crossings become anti-crossings. Already for the two-level system in a double quantum dot, one realizes that one actually has to deal with (dissipative) Landau-Zener tunneling between curves on energy surfaces in the parameter space of the problem [14]. This type of ‘pumping’, which occurs in Hilbert spaces that are essentially cut down to very small dimensions due to strong correlations, is the opposite limit of the ‘usual’ adiabatic pumping in large, non-interacting mesoscopic systems [15].

## 5 Conclusions

We have outlined the basic ideas of adiabatic passage and several examples of its application in solid-state devices that can be realized in superconductor as well as in semiconductor systems. The examples showed that adiabatic passage-like techniques may be used to generate simple charge transfers, to detect non-Abelian holonomies and to control charge transport properties (in particular also to design an alternative kind of charge pumps). In our opinion, it is obvious even from these simple examples that the method of adiabatic passage has significant potential for applications in solid-state physics and deserves appropriate attention also in this field.

The authors gratefully acknowledge stimulating discussions with L. Faoro, R. Fazio, A. Kuhn, F. Renzoni, and T. Vorrath.

## References

1. K. Bergmann and B.W. Shore, in: H.L. Dai and R.W. Field (Eds.) *Molecular dynamics and stimulated emission pumping*, p. 315 (World Scientific, Singapore, 1995).
2. K. Bergmann, H. Theuer, and B.W. Shore, Rev. Mod. Phys. **70**, 1003 (1998).
3. T. Brandes and F. Renzoni, Phys. Rev. Lett. **85**, 4148 (2000); T. Brandes, F. Renzoni, and R. H. Blick, Phys. Rev. B **64**, 035319 (2001).
4. F. Renzoni and T. Brandes, Phys. Rev. B **64**, 245301 (2001).
5. L. Faoro, J. Siewert, and R. Fazio, Phys. Rev. Lett. **90**, 028301 (2003).
6. M.H.S. Amin, A.Yu. Smirnov, and A. Maassen v.d. Brink, Phys. Rev. B **67**, 100508(R) (2003).
7. M.O. Scully and M.S. Zubairy, *Quantum Optics* (Cambridge Univ. Press, Cambridge 1997).
8. Yu. Makhlin, G. Schön, and A. Shnirman, Rev. Mod. Phys. **73**, 357 (2001).
9. A. Shapere and F. Wilczek (Eds.), *Geometric phases in physics*, (World Scientific, Singapore, 1989).
10. M.V. Berry, Proc. Roy. Soc. A **392**, 45 (1984).
11. F. Wilczek and A. Zee, Phys. Rev. Lett. **52**, 2111 (1984).
12. R.G. Unanyan, B.W. Shore, and K. Bergmann, Phys. Rev. A **59**, 2910 (1999).
13. L.-M. Duan, J.I. Cirac, and P. Zoller, Science **292**, 1695 (2001).
14. T. Brandes and T. Vorrath, Phys. Rev. B **66**, 075341 (2002).
15. P.W. Brouwer, Phys. Rev. B **58**, 10135 (1998).

# Quantum Dots: Building Blocks of Quantum Devices?

Manfred Bayer<sup>1</sup>, G. Ortner<sup>1</sup>, A. Larionov<sup>1</sup>, D.R. Yakovlev<sup>1</sup>, M. Schwab<sup>1</sup>,  
P. Borri<sup>1</sup>, W. Langbein<sup>1</sup>, U. Woggon<sup>1</sup>, I. Yugova<sup>1</sup>, G. Baldassarri Höger von  
Högersthal<sup>1</sup>, Y.B. Lyanda-Geller<sup>2</sup>, T.L. Reinecke<sup>2</sup>, S. Fafard<sup>3</sup>,  
Z. Wasilewski<sup>3</sup>, M. Korkusinski<sup>3</sup>, P. Hawrylak<sup>3</sup>,  
A. Forchel<sup>4</sup>, and J.P. Reithmaier<sup>4</sup>

<sup>1</sup> Experimentelle Physik II, Universität Dortmund  
44221 Dortmund, Germany

<sup>2</sup> Naval Research Laboratory  
Washington DC, 20375, USA

<sup>3</sup> Institute for Microstructural Sciences, National Research Council of Canada  
Ottawa, K1A 0R6, Canada

<sup>4</sup> Technische Physik, Universität Würzburg  
Am Hubland, 97074 Würzburg, Germany

**Abstract.** Among other systems quantum dots have been considered as one of the prime candidates for a solid state quantum information processing that is scalable up to a large number of quantum bits. After some general considerations we focus here on an essential building block of a quantum processor, a quantum gate for entangling the states of two quantum bits. A pair of coupled quantum dots could be used for such gating purposes. To that effect we study the suitability of vertically correlated InAs/GaAs quantum dot pairs fabricated by self-assembly. The exciton photoluminescence shows a characteristic splitting of the quantum dot shells, which hints at a tunnel coupling of the dots. This coupling has been confirmed by fine structure studies, in which distinct anticrossings are observed in the magnetic field dispersion. These avoided crossings would not occur if the two dots were decoupled. By applying an electric field along the molecule axis we demonstrate that the molecule coupling can be controlled, which is a prerequisite for an entanglement creating interaction of controllable strength between quantum bits. However, four-wave mixing studies show a strong shortening of the exciton dephasing in the molecules as compared to quantum dots which appears to be problematic for application of excitons as the genuine bits in quantum information processing. Thus excitonic complexes with their possibility to be controlled by ultrashort laser pulses may be useful only as intermediate states in schemes, where the quantum storage rests on spin degrees of freedom, but for manipulating spin is swapped into charge in form of charged excitons.

## 1 Introduction

During the last decade a lot of effort has been invested in developing semiconductor quantum structures, in which carriers are three-dimensionally confined [1]. The fabrication of these quantum dot systems has reached a remark-

able level of perfection. Motivation for these activities has come both from applications as well as from basic physics. One example has been the perspective to obtain optoelectronic devices such as lasers with an unprecedented performance, concerning threshold current, for example. Further, quantum dots might also allow to enter new wavelength ranges for light emitting devices.

In spectroscopic studies often a behavior is observed, which has strong similarities to that observed for atoms found in nature. It is this similarity, due to which quantum dots often are termed artificial atoms. It is also the reason, why quantum dots have attracted considerable attention from fields, which originated from atomic physics: quantum optics [2] and quantum information processing [3]. Their potential implementation in a solid state environment might offer the possibility to realize devices in which ultimately quantum effects are exploited.

One of the most prominent examples, for which a lot of progress has been made during recent years, is the single photon emitter [4]. Typically single photons are created by attenuation of intense laser pulses which for obvious reasons leads to a considerable uncertainty of the photon occupation number in the 'pulse' after attenuation. Here quantum dots appear to be very attractive, since pulsed injection of electron-hole pairs (either optically or electrically) in combination with spectral filtering of the emission offers the chance of efficient single photon creation. Indeed single photon generation and creation of non-classical light has been demonstrated now in a large number of experimental studies involving correlation spectroscopy. After these basic demonstrations, focus has shifted towards optimizing the performance of these emitters such as increasing operation temperature and emission efficiency.

But this might form only the start of the development of quantum optoelectronic solid state devices. Other examples are sources of entangled photons or squeezed light that could become important for lithography, microscopy, tomography or metrology. The development of corresponding devices still will require a lot of research. But also appliances which are common in everybody's life such as light emitting devices might benefit from quantum optics: One example is the precise control (enhancement or de-enhancement) of spontaneous emission of an optically active medium, that is placed in a resonator [5]. For this purpose various types of resonators which allow to confine light in different dimensionalities are being developed like photonic crystal defects, micropillars or microdisks.

Closely related to these problems from quantum optics are the tasks that are addressed in quantum information processing which comprises the fields of quantum cryptography, quantum teleportation or quantum computation all of which rely on encoding of information in form of quantum bits, which are two-level superspoititions  $\alpha|0\rangle + \beta|1\rangle$ . For example, in cryptography the information to be transmitted could be encoded in the polarization states of

a stream of single photons. For practical realization one obviously needs the aforementioned single photon sources.

Cryptography and teleportation rely on a modest number of quantum bits only, so that they may be more readily implemented than quantum computation, for which a large number of bits is required in order to solve tasks of practical relevance such as the factoring of large numbers or the search of large data bases. For the latter two examples it has been shown that quantum computers could solve these tasks with a speed that excels that of problem solution on a classical computer by orders of magnitude. This speed advantage relies on the entanglement of quantum bits, which simply speaking is the establishment of a controlled interaction among them, so that they are no longer individually addressed during computation, but a massive parallel processing becomes possible.

However, the road toward a quantum computer is extremely challenging as demonstrated by the five prerequisites defined by David di Vincenzo, which are shortly summarized here: well defined quantum bits have to be defined (1) which can be combined in large numbers (scalability). The quantum bit states can be initialized, for which they need to be decoupled (2). The decoherence time during which the quantum information is retained must be long enough to allow for coherent processing (3). Quantum gates must be realized which allow for entangling quantum bits and performing the parallel processing (4). And finally, after the processing has been done, it must be possible to read out the quantum result (5).

Many proposals for solid state quantum computing involve semiconductor quantum dots due to their similarity to two-level systems, other extremely prospective candidates might be superconductors. One of the biggest challenges in solid state systems is decoherence since coupling of the quantum processor to the environment beyond the degree necessary for manipulation purposes can hardly be avoided. Irrespective of the scheme to be implemented it seems clear already now that a quantum processor must be operated at cryogenic temperatures, otherwise thermal excitations will cause a too fast decoherence. We will also address the decoherence issue here, but mostly we will focus on point (4), the realization of an optically operated quantum gate based on quantum dots. Optical operation appears to be very attractive for proposals, in which presence and absence of an exciton (an electron-hole pair excitation) in a quantum dot defines the quantum bit. Such schemes could benefit from the possibility to perform coherent manipulation by ultrafast spectroscopy techniques on a sub-ps time scale, while the dephasing occurs in the ns-range [6]. Recent studies have demonstrated exciton Rabi-oscillations within a few ps [7]. But how could a gate be realized with excitons?

The simplest gating unit one can imagine is a pair of coupled quantum dots, a so-called quantum dot molecule. If the coupling strength can be controlled and in particular switched off, quantum bit excitons might be injected in each of the quantum dots independently. By re-establishing a coupling,

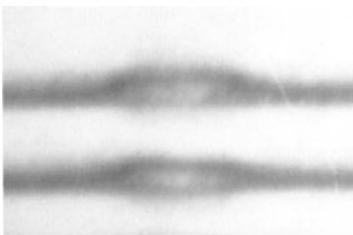
an entanglement of controllable strength might be created. Thus coupled quantum dots might form the basic building block of a quantum gate. For this purpose, the positions of the two dots relative to each other need to be controlled precisely. Here we focus on quantum dot molecules fabricated by self-assembly. We study the exciton states in these structures in detail and provide a spectroscopic proof for the tunnel coupling of the dots. We then aim at a control of the coupling between the dots through application of an electric field along the molecule axis. And finally we address the question whether the decoherence properties of the excitons are good enough to use them as the genuine quantum bits.

## 2 Coupled Quantum Dot Structures

Stransky – Krastanow growth seems ideally suited to create quantum dot molecules which fulfill the described requirements of a pair of correlated dot structures which shall serve as quantum gate. When growing two layers of dot structures in close vicinity, the strain that surrounds a quantum dot in a first, lower lying layer enforces the location of a second dot on top of the first one [8]. The relative positions are therefore well defined. An example of such a coupled dot structure is shown in the transmission electron micrograph in Fig. 1. Here the Stransky-Krastanow growth scheme was extended by a so-called Indium flush which allows for a shape engineering of the dots. When using the conventional scheme, the size of the dot in the second layer is considerably larger than that of the dot in the first layer which might aggravate a tunnel coupling of the dots. In order to make them as similar as possible, the In-flush was applied, by which the dot geometry is changed from a dome to a disk shape by razing off the top of the dome [9]. In this way also the penetration of carriers through the layer-shaped barrier between the dots is facilitated.

Samples with different barrier widths  $d$  were prepared, which were laterally patterned by lithography for isolating single molecules for spectroscopic purposes. The available barrier widths were 4, 5, 6, 7 and 8 nm as well as 16 nm as a reference for a double dot sample, for which tunnel coupling is negligible. Also a single quantum dot layer sample was available for reference purposes. Spectroscopic studies performed on single such dots demonstrate that the dots are to a good approximation rotationally invariant around the heterostructure growth axis.

From microscopy studies a high degree of double dot formation is found up to barrier widths of more than 10 nm. Whenever there is a dot in the first layer, there appears also a dot in the second layer. For the narrow barrier samples ( $d = 4$  or 5 nm) the vertical correlation of the two dot structures is very good, as exemplified in Fig. 1. No lateral shifts of the dot positions relative to each other are observed. This becomes different for the samples with larger barriers between the two dot layers, for which strain release aggravates



**Fig. 1.** Transmission electron micrograph of a self-assembled quantum dot molecule with a nominal InAs/GaAs material composition (courtesy of J.P. Mc Caffery, NRC Ottawa). The height of the dots is 1–2 nm, while their diameter is 15–20 nm. The width of the barrier between the quantum dots is 4 nm, as measured from wetting layer to wetting layer, so that the effective barrier width is 2–3 nm only [10]

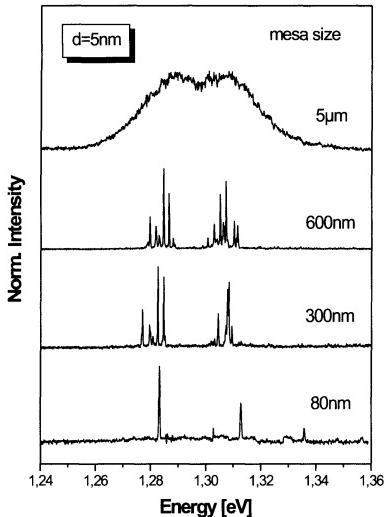
vertical dot correlation. For example for the samples with barriers of 7 and 8 nm a lateral displacement up to a few nm is observed for some double dot structures. For other structures still an almost perfect correlation of the dots is found. For the 16 nm barrier the displacement may be even larger.

### 3 Exciton States in Quantum Dot Molecules

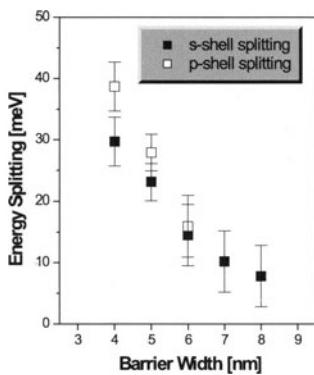
If the two quantum dots are coupled by tunnelling, a splitting of the quantum dot shells into bonding and antibonding orbitals is expected. Clear indications for that are found in photoluminescence studies of arrays of quantum dot molecules in a 5  $\mu\text{m}$  wide mesa structure (top trace in Fig. 2). A splitting of the emission corresponding to the s-shell in quantum dots by about 25 meV is observed [10]. This observation is consolidated by spectra taken on smaller mesa structures. For 600 and 300 nm wide structures, two emission bands consisting of several sharp emission lines are observed. When studying the smallest available mesa with a size of 80 nm, only two split spectral lines appear, which we attribute to emission from 'bonding' and 'antibonding' exciton molecule orbitals, respectively.

When performing such studies for the different molecule samples, we find a systematic increase of the splitting with decreasing barrier width [10,11]. The energy splittings of the s-shell and the p-shell versus barrier width are shown in Fig. 3. The data represent the values obtained by averaging many studies like those shown in Fig. 2. For narrow barriers large splittings of 30–40 meV are observed. The splittings for the p-shell are larger than those for the s-shell, since the p-shell states have a stronger penetration through the barrier, resulting in a larger tunnelling matrix element.

The picture of bonding and anti-bonding molecule orbitals is borrowed from a single particle model, while in our studies we investigate exciton states formed by an interacting electron-hole pair. Therefore it is necessary to take a closer look at these exciton states. For this purpose we have developed [10,11] a simple model which captures the essential physics of tunnel coupling, while neglecting complications that might arise from details of the molecule geometry or the semiconductor band structure, in particular that



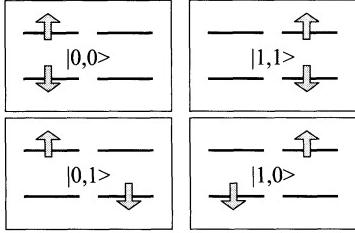
**Fig. 2.** Photoluminescence spectra at  $T = 2$  K of mesa structures with varying lateral sizes, that were prepared on an InAs/GaAs quantum dot molecule sample with a  $d = 5$  nm wide GaAs barrier [10]



**Fig. 3.** Barrier width dependence of the energy splittings between the exciton molecule shells, that arise from splitting of the s- and p-shells in the quantum dots [10]

of the valence band. For quantitative descriptions this model certainly has to be refined.

We start by neglecting Coulomb interaction: Electron and hole can be distributed in four different ways in the molecule, as shown in Fig. 4. They can be located on the same quantum dot (upper panels) or on opposite quantum dots (lower panels). When indexing the lower (upper) dot by '0' ('1'), these electron-hole states can be written as  $|0,0\rangle$  and  $|1,1\rangle$  for the first two configurations and as  $|0,1\rangle$  and  $|1,0\rangle$  for the other two configurations. Here the first (second) index gives the position of the electron (the hole). For being optically active, electron and hole have to be at the same position. Therefore  $|0,0\rangle$  and  $|1,1\rangle$  are optically active states, while  $|0,1\rangle$  and  $|1,0\rangle$  are optically inactive.



**Fig. 4.** The four different possibilities to distribute electron and hole in a double dot structure. Here we have indexed the lower quantum dot by '0', the upper quantum dot by '1'. The first (second) index in the state vectors gives the position of the electron (hole) [11]

Next the Coulomb interaction between electron and hole needs to be included. It leads to a mixing of the four single particle configurations. The resulting exciton states are to a good approximation given by

$$\Psi_1 \sim (|0,0\rangle + |1,1\rangle + |0,1\rangle + |1,0\rangle) \quad (1a)$$

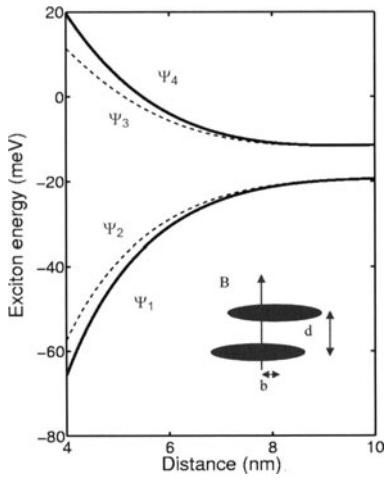
$$\Psi_2 \sim (|0,0\rangle - |1,1\rangle + |0,1\rangle - |1,0\rangle) \quad (1b)$$

$$\Psi_3 \sim (|0,0\rangle - |1,1\rangle - |0,1\rangle + |1,0\rangle) \quad (1c)$$

$$\Psi_4 \sim (|0,0\rangle + |1,1\rangle - |0,1\rangle - |1,0\rangle) \quad (1d)$$

Each electron-hole configuration contributes roughly with equal weight to the states. However when looking at their symmetry along the molecule axis under the restriction that electron and hole coordinates are equal,  $\Psi_1$  and  $\Psi_4$  turn out to be symmetric, while  $\Psi_2$  and  $\Psi_3$  are antisymmetric. Due to this symmetry only  $\Psi_1$  and  $\Psi_4$  are optically active, and it is these states which we observe in experiment. The two other states are dark because for calculating the oscillator strength the exciton amplitudes have to be summed, and the resulting transition matrix element vanishes for antisymmetric states. The calculated barrier width dependence of the energies of these four exciton states is shown in Fig. 5. From the data shown there we see, that the two optically active states indeed show a splitting behavior which follows very closely the intuitively expected one of bonding and anti-bonding orbitals.

The model developed above is for a molecule structure which consists of two identical quantum dots, an obviously unrealistic situation, since there will be always some inequality of the dots. In general, such an asymmetry will lead to a further enhancement of the splitting of the molecule states, which we demonstrate briefly for a single particle picture. For a symmetric quantum dot molecule the coupling of two electron states is described by a  $2 \times 2$  matrix  $\begin{pmatrix} E_e & t \\ t & E_e \end{pmatrix}$  where we have used the basis of delocalized states  $|0\rangle$  and  $|1\rangle$  for the matrix representation. Here  $E_e$  is the energy of the electron in a decoupled dot and  $t$  is the tunnelling matrix element. Diagonalizing this matrix by going from the basis of localized states to the one of delocalized (bonding and antibonding) states  $\sim (|0\rangle \pm |1\rangle)$ , gives the eigenenergies  $E_e \pm t$ . For an asymmetric structure the diagonal energies are different,  $E_{e,1} \neq E_{e,2}$ . Calculations show that the modification of the tunnelling matrix element by



**Fig. 5.** Energies of the four exciton states, that arise from quantum dot s-shells as functions of barrier width. Optically active states are shown by solid lines, while optically inactive states are shown by dotted lines [12]. The insert gives a sketch of a molecule structure, for which the two dots are displaced by  $b$

the asymmetry is quite small, as long as the asymmetry is not too large, as is the case for the structures under study. Diagonalization gives eigenenergies, whose splitting is always larger than that for the symmetric molecule structure

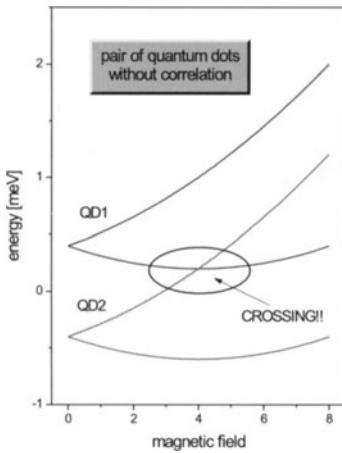
$$E_{1,2} = \frac{1}{2}(E_{e,1} - E_{e,2}) \pm \frac{1}{2}\sqrt{4t^2 - (E_{e,1} - E_{e,2})^2} \quad (2)$$

The asymmetry of the molecule structure will also lead to a mixing of the optically active and inactive states making the latter ones observable. To transfer a significant oscillator strength to  $\Psi_2$  and  $\Psi_3$ , a considerable asymmetry is required, for which we find for many of the molecule structures under study no indication at zero magnetic field. For completeness we note, however, that for small mesa structures we sometimes find emission spectra consisting of three or four lines. However, from photoluminescence studies it is not possible to assess the origin of these features in a single molecule.

## 4 Proof of Coherent Dot Coupling

While the spectroscopic data discussed above strongly support a tunnel coupling of the dot structures, they do not represent a unique proof. One might imagine, for example, that the strain surrounding the first dot varies systematically with increasing barrier width, so that the split emission features originate from decoupled quantum dots, whose emission energies vary systematically due to the systematic strain variation. Ideally, such a proof would be provided by corresponding studies on a single sample of given barrier width, and would not rely on comparative studies of samples with different barriers.

Such a tool was provided by studies of the fine structure of the exciton states confined in the quantum dot molecules [13]. Neglecting effects from the

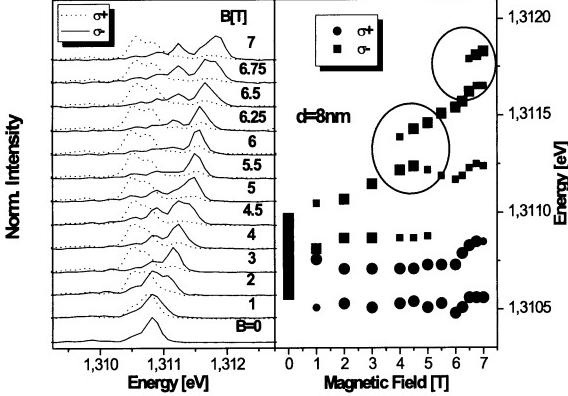


**Fig. 6.** Exciton fine structure pattern for two independent quantum dots. Here no mixing of dark and bright exciton spin states was assumed, so that the dark exciton energies are not shown

interaction with the lattice nuclei, this fine structure arises from the exchange interaction between electron and hole and couples their spins. The resulting total exciton angular momentum divides the states in those which can couple to the light field and those which are optically inactive. When applying a magnetic field, the fine structure is extended by the Zeeman interaction of the carrier spins with the field, which potentially gives more detailed insight due to the possibility to vary strength and orientation of the field.

Restricting to the ground state exciton in a GaAs-based quantum dot [14], the fine structure multiplett consists of four states: two with exciton angular momentum  $|M| = 1$  where  $M$  is the sum of the electron and hole angular momentum projections along the z-axis (the heterostructure growth axis). The other two states have angular momentum  $|M| = 2$ , and therefore they are not optically active, unless a strong symmetry breaking, which causes a mixing of bright and dark states, occurs. Such a symmetry breaking may be induced deliberately by applying a magnetic field with an orientation different from the symmetry axis of the quantum dots. It might, however, also occur for strong structural asymmetries.

After these considerations we can turn to the exciton fine structure in quantum dot molecules. Before we present the experimental data, we discuss the fine structure pattern that we would expect in case that the two dots are not coupled. These expectations are given in Fig. 6, where we restrict to the energy range of the ground state excitons  $\Psi_1$  and  $\Psi_2$  in Fig. 5. Since the dots cannot be made identical on an atomistic level, at  $B = 0$  two emission lines would appear in the spectra. Each of these lines will show a doublet spin-splitting in magnetic field, assuming that the dark exciton states do not mix with the bright ones. When features from the two doublets are brought into resonance by ramping  $B$ , they cross each other without interaction because of the decoupled nature of the quantum dots.

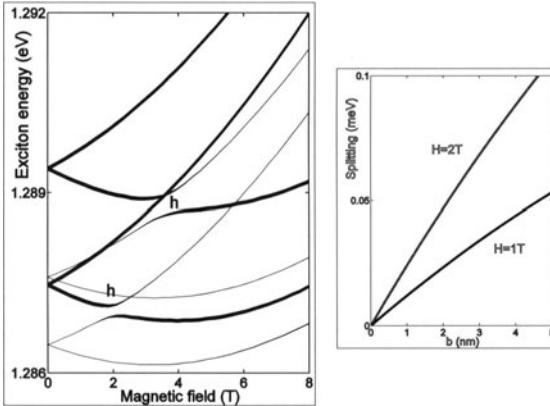


**Fig. 7.** The left panel shows photoluminescence spectra of a single quantum dot molecule with a  $d = 8$  nm wide barrier for different magnetic fields applied along the molecule axis. The circular polarization of the emission was analyzed. The right panel shows the magnetic field dispersion of the exciton transition energies extracted from the spectra in the left panel. The symbol sizes give the relative intensities of the transitions, the circles indicate avoided crossings [10]

The left panel of Fig. 7 shows photoluminescence spectra of a quantum dot molecule with a barrier width of  $d = 8$  nm, recorded at different magnetic fields. The field was oriented along the molecule axis (the Faraday-configuration). At  $B = 0$  a single emission line is observed, which arises from exciton recombination out of state  $\Psi 1$ . In magnetic field the line splits due to the Zeeman interaction. For intermediate fields clearly five spectral lines are observed. This demonstrates that not only a single quantum dot is involved in the emission. For this a maximum of four emission lines could occur, as discussed above. Instead, at least two quantum dots must be involved.

More insight is obtained from the magnetic field dispersion of the emission features. At  $B = 4$  T the high energy side of the emission is dominated by a  $\sigma-$ -polarized spectral line. At  $B = 4.5$  T another spectral feature appears at higher energies. When increasing the field strength further, the low energy line fades away and the oscillator strength is transferred to the high energy line. During this exchange process, the energy of the lower energy line, which increases with  $B$  below 4 T, saturates while the emerging line starts to shift to higher energies. These observations are clear indications for an avoided crossing between the two lines. At higher magnetic fields another such avoided crossing appears in the spectra.

The behavior becomes more obvious in the magnetic field dispersion of the exciton transition energies, which is shown in the right panel of Fig. 7 for this quantum dot molecule. The size of the symbols gives the relative emission intensity of the different features. Clearly two anticrossings processes are observed around 4.5 T and 7 T, as highlighted by the circles. The appearance

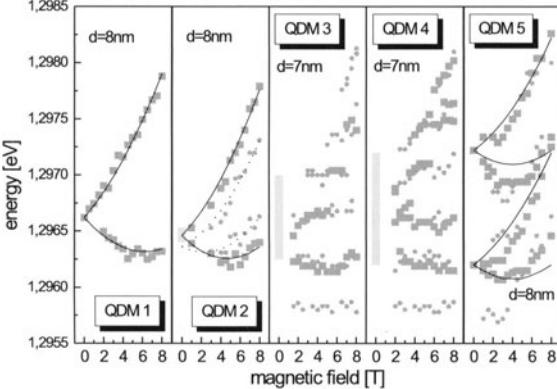


**Fig. 8.** Left panel: Magnetic field dispersion of the exciton fine structure of states  $|\Psi_1\rangle$  and  $|\Psi_2\rangle$  for an asymmetric quantum dot molecule (see text). Right panel: Magnitude of the avoided crossings as function of lateral dot displacement for two different magnetic field strengths [12]

of these anticrossings is a clear proof for the quantum mechanically coherent coupling of the two quantum dots. They could not occur for decoupled dots.

Finally the origin of the anticrossings has to be clarified, for which we need mechanisms which in effect lead to spin-flip processes so that different angular momentum states are mixed. For this purpose a detailed theoretical modelling has been performed [13,12], from which we find that the physical origin for the holes lies in an anisotropic Zeeman interaction: It arises from mixing of the  $\Gamma_8$  and  $\Gamma_7$  valence bands in the Kane-Hamiltonian and acts like an in-plane magnetic field. For the electrons it arises from intraband spin-orbit coupling.

To see the consequences of these interactions, let us start by considering an ideal molecule structure consisting of two identical quantum dots that are not laterally displaced with respect to each other. Each of the exciton states  $\Psi_i$ ,  $i = 1, \dots, 4$ , forms a fine structure multiplett consisting of four exciton states, corresponding to the possible spin orientations of electron and hole, very much like for an exciton in a single quantum dot. As discussed above, we restrict to the energy range, in which the bright state  $\Psi_1$  and the dark state  $\Psi_2$  are located (see Fig. 5). Here bright and dark refer to the optical properties that arise from the symmetry of the exciton wave function along the molecule axis. In the spectra we would observe then a doublet splitting of the spin-bright states of  $\Psi_1$  with exciton angular momenta  $|M| = 1$ , while the spin dark states with  $|M| = 2$  would remain dark. Such a molecule would therefore show the same fine structure splitting as an exciton in a quantum dot, so that no conclusive result about the molecule formation can be obtained from such investigations.



**Fig. 9.** Magnetic field dispersions of the exciton fine structure for 5 additional InAs/GaAs quantum dot molecules with barrier widths  $d$  of 7 or 8 nm [13]

However, for the molecules with rather wide barriers ( $d = 7$  or  $8$  nm) we often find a lateral displacement of the dots, as discussed in section 2. From our modelling we find that it is these asymmetries which give rise to fine structure patterns like the one in Fig. 7 because they switch on the physical mechanism for mixing spin states as discussed above. Figure 8 shows an example of the effects of asymmetry on the fine structure. Here we have assumed the following dot properties: the confinement potentials for electrons and holes are different, for the electrons by 20 meV, for the holes by 3 meV. The dot radii differ by 0.2 nm assuming 8.0 nm for dot 1 and 8.2 nm for dot 2. Further the two dots are laterally displaced by 4.0 nm. Due to this asymmetry a mixing of  $\Psi_1$  and  $\Psi_2$  occurs, in particular when a magnetic field is applied, making the  $\Psi_2$ -states optically active. Since each state consists of four different spin states, up to 8 emission lines might appear in the spectra, giving an explanation for the increased number of spectral features in Fig. 7.

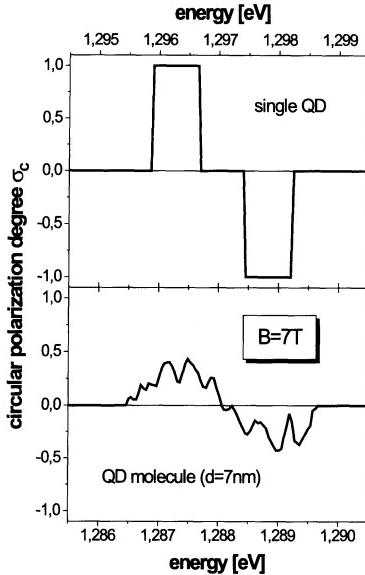
The calculated magnetic field dispersion of the exciton fine structure of states  $\Psi_1$  and  $\Psi_2$  is shown in the left panel of Fig. 8 and shows indeed quite a complicated behaviour with two pronounced anticrossings labelled by h, indicating that their origin lies in the valence band holes. Note that we did not aim for a quantitative agreement between experiment and theory since too many parameters of the quantum dot molecules are not known with high enough accuracy. However we want to stress again, that the anticrossings can occur only if the two dots are tunnel coupled, both for electrons and holes. The dependence of the magnitudes of the level splittings due to avoided crossings is shown in the right panel of Fig. 8. The splittings increase with increasing  $B$ , since by doing so also the effective in-plane magnetic field is increased, promoting spin flip processes and thus level mixings. They also increase with increasing lateral displacement of the dots corresponding to larger asymmetry of the molecule structure.

We have studied a large number of quantum dot molecules. Examples of the observed magnetic field dispersions of fine structure patterns are given in Fig. 9. From these studies we find strong variations from molecule to molecule. By comparison with the calculations we can trace this behavior to variations of geometry and symmetry of the structures, which depend on the particular structure under study. The spectrum we find for these wide barrier samples ranges from high to low symmetry.

For high symmetries we find – in accordance with the theoretical expectations - an exciton doublet splitting (QDM1), in some cases also a quadruplet splitting (QDM2). Since similar patterns are also observed for quantum dots, this result cannot be used as demonstration of molecule formation. To this end, measurements of the exciton diamagnetic shift in the Voigt-configuration (normal to the heterostructure growth direction) have been performed (not shown), which allow to estimate the extension of the exciton wave function along the molecule axis [13,12]. For the molecules we find a shift that is about three times larger than that observed for the dots indicating that the exciton wave function extends across the two dots and the barrier.

For intermediate symmetries (QDM3 and QDM4) we find anticrossing processes as for the molecule structure discussed above. For low symmetries (QDM5) we find a behavior which at first sight looks like the emission of two independent quantum dots. These data have some similarity with the calculated splitting pattern for the molecule structure in Fig. 8. At  $B = 0$  two spectral lines appear which show a doublet splitting in magnetic field. The available magnetic fields were not large enough to bring features from the two doublets into resonance and look for anticrossings for judging about coherent coupling. However, there are a couple of observations which set QDM5 clearly apart from decoupled dots. First, for the two doublets at low  $B$  negative diamagnetic shifts are observed. Second, the spin-splitting of the two doublets depends in a highly non-linear fashion on magnetic field. Both results have never been observed for quantum dots and are also not expected for them. While a detailed explanation is missing, the negative diamagnetic shift might arise from a redistribution of the exciton wave function in the molecule by the magnetic field, as has been observed for trions in higher dimensional semiconductors.

Reductions of the symmetry of the quantum dot molecules have also drastic consequences for the polarization of their emission [10]. Quantum dots suffer often from a linear distortion of their geometry, resulting in a linear polarization of the emission at zero magnetic field. However, in this case already moderate magnetic fields of about 1 T are sufficient to restore the cylindrical symmetry and to induce a circular polarization. This is demonstrated in the upper panel of Fig. 10, which shows that the circular polarization degree of the emission of a quantum dot at  $B = 7$  T is fully circularly polarized,  $\sigma+$  polarized for the low energy line,  $\sigma-$  polarized for the high energy line. Here the polarization degree is defined as the difference between



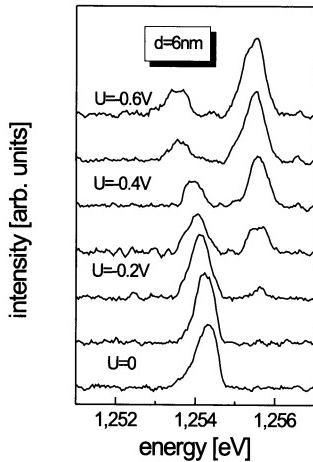
**Fig. 10.** Circular polarization degree of the emission as function of energy for a single quantum dot (upper panel) and for a quantum dot molecule with a  $d = 7$  nm wide barrier. The data have been taken at a magnetic field strength of 7 T [10]

$\sigma+$  and  $\sigma-$ -polarized emission intensity devided by the sum. The reason for the restoration of circular polarization is the dominance of the Zeeman interaction over the electron-hole exchange coupling which occurs already at small fields of a few T.

The quantum dot behavior has to be contrasted with the one observed for a quantum dot molecule with a 7 nm barrier (QDM4 in Fig. 9), whose fine structure multiplett shows a number of avoided crossings. The corresponding data are given in the lower panel of Fig. 10, showing that even high magnetic fields are not sufficient to restore the cylindrical symmetry. The maximum polarization degree that can be obtained is about 40%. It varies almost continuously over the emission band, in contrast to the almost digital behavior observed for the quantum dot. These results aggravate proposals in which tunnel coupled quantum dots serve as source for polarization entangled photon pairs [15].

For completeness we mention also the fine structure data for narrow barrier samples. For them, the energy splitting between  $\Psi_1$  and  $\Psi_2$  has increased so much (see Fig. 5), that in the magnetic field range studied here, fine structure levels from these two doublets cannot be brought into resonance. A single ground state emission line can therefore split at most into four features, for most structures we observe a doublet splitting.

Finally, we also note that there are a large number of other spectroscopic quantities all of which show a systematic dependence on the width of the barrier between the quantum dots [10]. They thus all clearly hint at coherent coupling of the dot structures. Without showing them here, examples for such quantities are [10]: the diamagnetic shift of the exciton when the magnetic



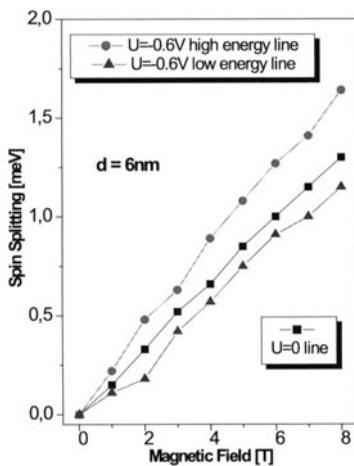
**Fig. 11.** Photoluminescence spectra of a self-assembled  $In_{0.6}Ga_{0.4}As/GaAs$  quantum dot molecule with a 6 nm wide barrier embedded in a diode structure for different voltages applied in reverse bias [16]

field is applied along the molecule axis (the Faraday-configuration); the exciton spin-splitting in this configuration; all the quantities needed to describe the exciton dephasing in the molecules (see next section) such as dephasing time, zero-phonon weight etc.

## 5 Electric Field Control of Molecule Bonding

Irrespective of the ultimate scheme to be implemented for quantum information processing, for obtaining a controlled entanglement an interaction between the quantum bits needs to be established, whose strength can be controlled on a detailed level. We want to demonstrate that for the quantum dot molecules studied here such a control can be achieved by applying an electric field along the molecule axis [16]. For this purpose,  $In_{0.6}Ga_{0.4}As/GaAs$  quantum dot molecules were placed in a diode structure. Fig. 11 shows photoluminescence spectra of a single such molecule for varying biases applied in reverse direction, so that leakage currents are suppressed. Again we focus on the energy range in which the exciton states  $\Psi_1$  and  $\Psi_2$  are located. In the flat band situation ( $U = 0$ ) a single emission line due to the radiative recombination of  $\Psi_1$  is observed. With increasing bias this line shows the well known Stark shift and its intensity is reduced. Simultaneously a new emission line appears on the high energy side with an energy which depends only weakly on bias, while its intensity increases strongly so that it becomes the dominant line.

It might be argued, that the spectral feature appearing on the high energy side could arise from charged exciton recombination. To exclude such an interpretation, we have performed a number of experimental tests. The first one was the measurement of the exciton spin-splitting in a magnetic field applied along the molecule axis. For charged and for neutral exciton one expects the same splitting, since the underlying g-factors are given by those



**Fig. 12.** Magnetic field dependence of the spin-splitting of the emissions lines observed for a  $d = 6$  nm quantum dot molecules at  $U = 0$  and at  $U = -0.6$  V

of the recombining electron-hole pair. These g-factors are identical for exciton and trion. The spin splittings that we observe for the line at  $U = 0$  and for the two lines at  $U = -0.6$  V are shown in Fig. 12 versus magnetic field. Within the experimental accuracy the splittings depend linearly on magnetic field. Most importantly, we find at  $U = -0.6$  V for the low energy line a splitting of 1.1 meV, while the high energy line shows a much larger splitting of 1.7 meV, preventing the charged exciton interpretation.

Measurements of the diamagnetic shifts of the lines validate this, and further, they give some insight into the spatial distribution of the wave function to be discussed in the following. In these studies (not shown) we find that the shift of the single emission line observed for flat band situation is the mean value of the shifts of the two features in the tilted band structure situation. This finding is within a few percent in excellent agreement with the results of detailed calculations.

The application of the electric field leads to a strong redistribution of the exciton wave function in the molecules. For  $U = 0$  the excitons are extended over the whole molecule structure. When ramping  $U$ , the behavior of the ground state  $\Psi_1$  follows the naïve expectations, that electron and hole are separated and become located on opposite dots. Thus its character changes into that of an exciton with dominant interdot character, which can be written as  $\Psi_1 \rightarrow |1,0\rangle$  by using the indices given above to the dots. Due to the spatial separation of electron and hole the oscillator strength of  $\Psi_1$  is strongly reduced. The first excited state  $\Psi_2$  undergoes as well a strong spatial redistribution. In contrast to  $\Psi_1$ , it is transformed into a state with dominant intradot exciton character, and thus becomes optically active, as seen experimentally  $\Psi_2 \rightarrow |0,0\rangle$ . For completeness, we note that state  $\Psi_3$  undergoes a similar transformation as  $\Psi_2$ , but electron and hole will become

located in dot 1 ( $\Psi_3 \rightarrow |1, 1\rangle$ ).  $\Psi_4$ , on the other hand, becomes an optically inactive state because  $\Psi_4 \rightarrow |0, 1\rangle$ .

When studying molecule structures with different barrier widths, we find that the voltage that has to be applied to induce the transformation from extended to inter- or intradot exciton states, is the larger, the narrower the barrier between the dots is (not shown). This finding is well conceivable, since for narrower barriers the coupling of the dots increases as evidenced by the enhancement of the tunnel splitting and thus becomes more rigid. Therefore a larger tilting of the band structure by the electric field is required, in order to cause the redistributions of the exciton wave function described before.

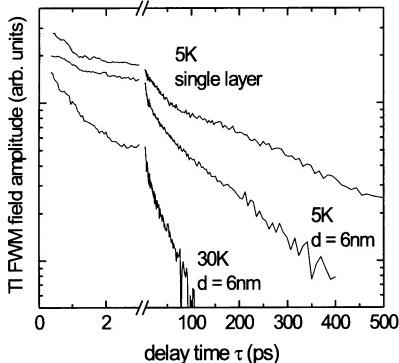
Based on this control of the electron-hole pair states in space, an interaction of controllable strength among quantum bits may be established. Such an interaction could be the dipole interaction among two excitons, if the bits are to be defined by excitons. Or for electrons and their spin degrees of freedom in quantum bit applications, this interaction would be the electron-electron exchange interaction, whose strength might be as well varied over wide ranges by the demonstrated electric field control.

## 6 Exciton Dephasing

After having obtained insight into the exciton states in the quantum dot molecules, we now turn the analysis of their dephasing properties. The results of corresponding four-wave mixing studies [17] at cryogenic temperatures are shown in Fig. 13. The dephasing occurs on two different time scales, a short one of a few ps and a long one of hundreds of ps, as shown by Borri et al. [6]. The short time dephasing is due to the dressing of electron-hole pairs by phonons, and gives rise to a broad phonon background. The long time dephasing at  $T = 5\text{ K}$ , where the phonons are basically frozen out, is determined by the lifetime of the excitons in the ground state. In these strongly confined systems excitons cannot be scattered into other quantum states, so that the lifetime should be the radiative decay time, as confirmed in quantum dot studies. In the spectrum it gives rise to the zero phonon line.

Here we are primarily interested in the long delay time behavior. For a single layer of quantum dots an exciton dephasing time of about 600 ps at  $T = 5\text{ K}$  is observed, corresponding to a homogeneous line width of  $\sim 2\mu\text{eV}$ . This is in good agreement with data reported in other studies. Turning to quantum dot molecules with a barrier width of  $d = 6\text{ nm}$ , a significant shortening of the dephasing time at  $T = 5\text{ K}$  down to 300 ps (equivalent to a homogeneous line width of  $4\mu\text{eV}$ ) is observed. When increasing the temperature to  $30\text{ K}$ , further shortening occurs.

Fig. 14 gives the barrier width dependence of the homogeneous line width at  $T = 5\text{ K}$ . With decreasing  $d$  a systematic increase of the homogeneous line width is observed. Most importantly, for the two samples with the narrowest



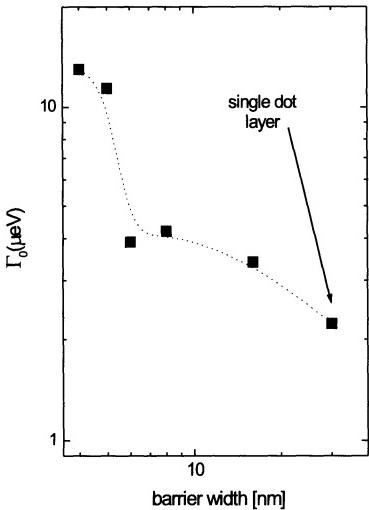
**Fig. 13.** Four wave mixing traces recorded on a reference single dot layer sample at  $T = 5\text{ K}$  (top trace) and a quantum dot molecule sample with a  $d = 6\text{ nm}$  barrier at  $5\text{ K}$  and  $30\text{ K}$  (the other two traces) [17]

barriers it is more than  $10\text{ }\mu\text{eV}$ , corresponding to dephasing times of slightly more than 100 ps only. This is equivalent to a decrease of the dephasing time by a factor of about 5. What is the origin of this decrease? As discussed above, for quantum dots the dephasing is determined by the radiative decay of excitons at low temperatures. If one makes this assumption also for the molecules, one has to consider the modification of the coherence volume of the exciton by coupling two quantum dots. The coherence volume determines the exciton oscillator strength and therefore also its radiative life time.

Assuming that the exciton wave function is extended over the molecule structure, its coherence volume is increased by a factor of about 2, taking into account also its penetration into the barrier. This increase translates into a reduction of the exciton life time and a corresponding increase of the line width. However, this effect will make for a factor of 2, not a factor of 5, as observed experimentally. In recent population lifetime measurements indeed a reduction of the life time by a factor of 2.5 was observed, confirming the simple estimate regarding the coherence volume of the exciton.

Currently it is not clear where the much stronger increase of the line width as expected from the radiative decay comes from. It might be possible that - in contrast to quantum dots, where at low temperatures no pure dephasing occurs – for the quantum dot molecules pure dephasing channels open up which lead to the much more drastic dephasing enhancement. Finally let us discuss the impact of these results on quantum information processing application in which the exciton in a quantum dot molecule shall be used as the genuine quantum bit. The obtained results certainly represent a major obstacle for such proposals. Recent works on exciton Rabi-oscillations [7] demonstrate that single exciton quantum bit rotations can be performed on a ps time scale. Comparing this time scale to the dephasing time which sets the upper limit for coherent manipulations, it seems hardly possible to realize a quantum computing device, in which exciton quantum bits are used also for storage purposes.

Therefore other approaches have to be taken. Recent studies have shown that the electron spin relaxation time is much longer than the exciton spin



**Fig. 14.** Dependence of the homogeneous line width at  $T = 5$  K as function of the barrier width between the two dots that form the molecule [17]

flip time. For gated quantum dots prepared on two-dimensional electron gases spin relaxation times exceeding  $50 \mu\text{s}$  have been observed [18]. Corresponding studies on self-assembled quantum dots still need to be performed, but calculations show that the relaxation time should be in the  $\mu\text{s}$ -range. The limiting factor is the hyperfine interaction of the electron spins with the lattice nuclei [19]. Therefore it appears to be appropriate to use the electron spin degree of freedom for quantum storage of information.

In order to exploit the experimental possibilities for coherent manipulation that are offered by ultrafast spectroscopy, these spin quantum bits need to be converted into charge. This can be done by 'adding' and 'subtracting' an electron-hole pair to the electron spin (thus forming an intermediate state charged exciton) by coherent Rabi rotations with suitably tailored laser pulses. It has been shown, that by exploiting the dipole interaction between trions in adjacent quantum dots a phase gate can be constructed [20]. In these proposals the charged exciton is therefore not used for storage, but only for coherent manipulation, while the storage is taken over by the spin. A 100 ps decoherence time seems long enough for performing a sufficient number of coherent manipulations in form of Rabi-oscillations. This seems an attractive pathway which would allow for an optical quantum information processing involving semiconductor quantum dots.

In this respect, also another result of the four-wave mixing studies is encouraging. As mentioned at the beginning of this section, the spectral line is composed of the zero phonon line superimposed on a broad background due to exciton-phonon interaction. In coherent control experiments the excitation of lattice vibrations as origin of decoherence should be avoided. In this respect the broad phonon background would represent an obstacle. In the quantum dots the zero phonon line contributes 75% to the signal, the rest stems from

the background and is therefore quite significant. For the molecules we find a considerable reduction of the phonon background. For the narrow barrier samples its importance is down to 10% , so that the zero phonon line can be addressed more easily.

## 7 Summary

In summary, we have tried to develop a detailde picture of the exciton states in self-assembled quantum dot molecules. Based on the entirety of the obtained results, it has been clearly demonstrated that the two dots are quantum mechanically coherently coupled, although in particular the role of the hole needs to be understood in much more detail. To obtain such an understanding seems difficult in experiments with interband spectroscopy by which always the combination of electron and hole properties ia addressed. In case of some asymmetry of the molecule, which will be reflected by the exciton wave function, electric field control might turn out as the solution, since it could be used to compensate for asymmetry.

For quantum information processes, most of the required demonstrations by cw-spectroscopy have been performed on the molecules and the stage is set to push hard for time-resolved spectroscopy to reach coherent control. Here during recent years a lot of progress has been made on so-called interface or natural quantum dots with their rather large sizes and small confinement potential height [21]. Despite of intense efforts, the level that has been reached for self-assembled dots is currently far from that on the interface dots. The origin for this is the exciton oscillator strength which is up to an order of magnitude smaller for the self-assembled structures. To demonstrate similar quantum control on them, spectroscopic techniques need to be improved further, both on the side of the equipment (light sources and detectors) and also on the side of applied techniques.

## Acknowledgements

We gratefully acknowledge the financial support of this work by the Deutsche Forschungsgemeinschaft (Forschergruppe 'Quantum optics in semiconductor nanostructures'), by the Quist program of DARPA and by the IST-FET program CECQDM of the European Commission. We thank A. Kress for his participation in the electric field studies.

## References

1. M. Grundmann (ed.), *Nano-Optoelectronics concepts, physics and devices*, (Springer-Verlag, Berlin 2002); L. Jacak, P. Hawrylak, and A. Wojs, *Quantum Dots*, (Springer Verlag, Berlin, 1998); M. Grundmann, D. Bimberg, and N.N. Ledentsov, *Quantum Dot Heterostructures*, (Wiley, New York 1998).

2. P. Berman (ed.), *Cavity Quantum Electrodynamics*, (Academic Press, San Diego, 1994).
3. D. Bouwmeester, A. Ekert, and A. Zeilinger *The Physics of Quantum Information* (Springer, Berlin 2000).
4. J. Kim et al., Nature **397**, 500 (1999); P. Michler et al., Science **290**, 2282 (2000); C. Santori et al., Phys. Rev. Lett. **86**, 1502 (2001); Z. Yuan et al., Science **285**, 102 (2001); M. Pelton et al., Phys. Rev. Lett. **89**, 233602 (2002).
5. E.M. Purcell, Phys. Rev. **69**, 681 (1946); J.M. Gerard et al., Phys. Rev. Lett. **81**, 1110 (1998).
6. P. Borri et al., Phys. Rev. Lett. **87**, 157401 (2001); D. Birkedal et al., Phys. Rev. Lett. **87**, 227401 (2001); M. Bayer et al., Phys. Rev. B **65**, 041308(R) (2002); C. Kammerer et al., Phys. Rev. B **66**, 041306(R) (2002).
7. H. Kamada et al., Phys. Rev. Lett. **87**, 246401 (2001); H. Htoon et al., Phys. Rev. Lett. **88**, 087401 (2002); A. Zrenner et al., Nature **418**, 612 (2002); P. Borri et al., Phys. Rev. B **66**, 081306(R) (2002).
8. Q. Xie et al., Phys. Rev. Lett. **75**, 2542 (1995); G.S. Solomon et al., Phys. Rev. Lett. **76**, 952 (1996).
9. S. Fafard et al., Phys. Rev. B **59**, 15368 (1999); S. Fafard et al., Appl. Phys. Lett. **76**, 2268 (2000).
10. G. Ortner et al., submitted for publication.
11. M. Bayer et al., Science **289**, 451 (2001).
12. Y. B. Lyanda-Geller et al., submitted for publication.
13. G. Ortner et al., Phys. Rev. Lett. **90**, 086404 (2003).
14. see M. Bayer et al., Phys. Rev. B **65**, 195315 (2002) and references therein.
15. O. Gywat et al., Phys. Rev. B **65**, 205329 (2002).
16. G. Ortner et al., submitted for publication.
17. P. Borri et al., Phys. Rev. Lett. **91**, 267401 (2003).
18. R. Hanson et al., Phys. Rev. Lett. **91**, 196802 (2003).
19. L.M. Woods et al., Phys. Rev. B **66**, 161318 (2002).
20. T. Calarco et al., Phys. Rev. A **68**, 012310 (2003); E. Pazy et al., Europhys. Lett. **62**, 175 (2003).
21. see, for example, T.H. Stievater et al., Phys. Rev. Lett. **87**, 133603 (2001); Gang Chen et al., Phys. Rev. Lett. **88**, 117901 (2002); X.Q. Li, Science **301**, 809 (2003); T. Guenther et al., Phys. Rev. Lett. **89**, 057401 (2002).

# The Structure and Dispersion of a Sharp Quantum Hall Edge Probed by Momentum-Resolved Tunneling

M. Huber<sup>1</sup>, M. Grayson<sup>1</sup>, M. Rother<sup>1</sup>, W. Biberacher<sup>2</sup>, W. Wegscheider<sup>1,3</sup>,  
and G. Abstreiter<sup>1</sup>

<sup>1</sup> Walter Schottky Institut, Technical Universität München  
85748 Garching, Germany

<sup>2</sup> Walther-Meissner-Institut, BAMW München  
85748 Garching, Germany

<sup>3</sup> Universität Regensburg  
93040 Regensburg, Germany

**Abstract.** We present measurements of momentum-resolved magneto-tunneling from a perpendicular two-dimensional (2D) contact into integer quantum Hall (QH) edges at a sharp edge potential created by cleaved edge overgrowth. Resonances in the tunnel conductance correspond to coincidences of electronic states of the QH edge and the 2D contact in energy-momentum space. We directly probe the dispersion curves of individual integer QH edge modes, and find that an epitaxially overgrown cleaved edge realizes the sharp edge limit, where the Chklovskii picture relevant for soft etched or gated edges is no longer valid. The inter-channel distances turn out to be smaller than both the magnetic length and the Bohr radius. We provide a quantitative model for describing the lineshape of conductance peaks in quantum Hall edge tunneling. The zero bias tunnel conductance in our single edge tunnel experiment does not feature the persistent conductance previously reported for a double quantum Hall edge tunnel geometry, and we explain this as a consequence of using a probe contact with a non-fluctuating Fermi energy. Furthermore with the measured dispersion relation reflecting the potential distribution at the edge we directly deduce the band bending at our cleaved edge under the influence of an external voltage bias. At finite bias we observe significant deviations from the flat-band condition in agreement with self-consistent calculations of the edge potential.

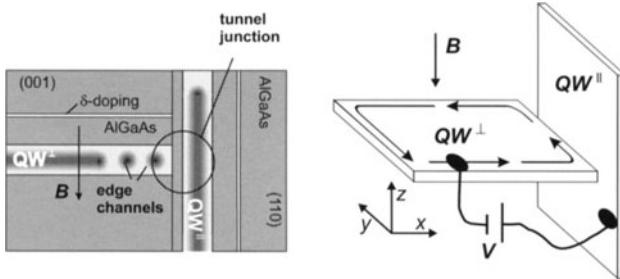
## 1 Introduction

The quantum Hall (QH) effect arises due to energy gaps developing in the spectrum of two-dimensional electron systems with a perpendicular magnetic field. At specific ratios of electron sheet density and magnetic field, when the Fermi-energy is in such a gap, the only low-energy excitations in a finite quantum Hall sample are located at the boundary, where the energetically bent Landau-levels intersect the Fermi-energy. These states at the sample edge exhibit one-dimensional (1D) chiral transport behavior, i.e. quantized conductance, independent of the exact electrostatics at the sample boundary,

so standard transport measurements do not provide information about such physics as depletion lengths, edge reconstructions, or edge excitations. The technique of edge tunneling in cleaved-edge overgrowth devices was introduced to address the last question of edge excitations in particular. 2D-3D tunneling at a sharp edge was used to observe the power-law density of states of fractional quantum Hall edges [1,2,3,4] and discrepancies among experiments suggested that the exact potential shape at the edge might be crucial for defining the correlations at the QH edge. By replacing the 3D disordered contact with a high mobility 2D system, one can perform momentum resolved tunnel spectroscopy of the QH edge.

Momentum-resolved tunneling has been used to measure both the dispersion relation of electronic excitations, as well as their momentum-resolved density of states or spectral functions. Spectral functions of two-dimensional (2D) [5] and one-dimensional (1D) [6] systems have been experimentally measured, and the dispersion relations of 1D systems have shown evidence for spin-charge separation [7]. In the quantum Hall regime, Kang, et al. [8] fabricated a pair of coplanar 2D systems laterally coupled through a tunnel barrier, and were able to map out the dispersion of the integer QH edge by probing one QH system with another. Additional theoretical works [9,10,11,12] have undertaken to model these and other systems, and discuss how dispersion relations and spectral features should be reflected in magnetotunneling conductances between the various 1D and 2D combined systems.

Focusing on the quantum Hall edge case, several outstanding puzzles remain in the results of Kang, et al. [8]. The use of a sharp edge model was shown to be consistent with the observed data, but no real-space distance scales were measurable in this experiment to quantify this assumption. Although the qualitative fit to this sharp edge was good, the conductance peaks did not occur at the expected energy value, nor was an unanticipated broad persistence of the zero-bias conductance peak over a wide range of filling factor explained. Theoretical descriptions of this device which included interaction effects still underestimated the width of this peak by almost an order of magnitude [11,12,13]. Kang, et al. also pointed out a curious absence of spin-splitting in the observed spectrum. Finally, no data was presented on the lineshape of the zero-bias conductance peak with magnetic field sweep, which could give insight into the spectral functions at the QH edge. This Letter is addresses all of these questions by implementing a different tunneling geometry consisting of two *orthogonal* quantum wells. Similar to Kang, et al. we see no evidence of spin-splitting, but in contrast we see no anomalously broad zero-bias peak, a fact we attribute to using a probe system with a non-fluctuating ground energy. In our geometry we can measure the real-space positions of the edge states, and we present tunnel conductance data which describes the spectral functions in the system. Furthermore, applying a finite voltage bias we study the edge dispersion of Landau levels in the integer QH



**Fig. 1.** The samples are fabricated by cleaved edge overgrowth. Two quantum wells ( $QW^\perp$  and  $QW^{\parallel}$ ) are arranged in a T-shape separated by a 50 Å thick tunnel barrier. A magnetic field  $B$  creates quantum Hall edge currents in  $QW^\perp$  propagating parallel to the extended tunnel junction.  $QW^{\parallel}$  acts as probe contact

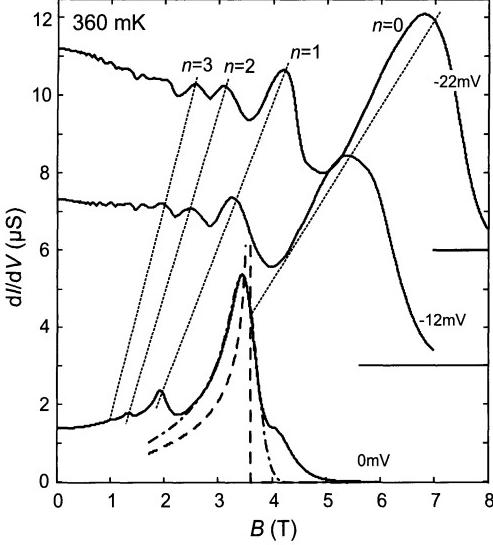
regime and address for the first time the effect of edge electrostatics on the observed resonance positions.

## 2 Sample Design

The samples consist of two separately contacted perpendicular high mobility quantum wells ( $QW^\perp$  and  $QW^{\parallel}$ ) forming a T-shaped structure (Fig. 1), where  $\perp$  and  $\parallel$  are defined relative to the quantizing  $B$  field which identifies  $QW^\perp$  as the quantum Hall effect system under study, and  $QW^{\parallel}$  is therefore referred to as the probe quantum well (Fig. 1b). The QWs consist of GaAs embedded in  $Al_{0.32}Ga_{0.78}As$ . Using cleaved-edge overgrowth (CEO), a 150 Å thick (001)-quantum well ( $QW^\perp$ ) is cleaved along the perpendicular (110)-plane and overgrown with a  $w = 200$  Å thick (110)-quantum well ( $QW^{\parallel}$ ) in a second epitaxial growth step. The quantum wells are separated from each other by a  $b = 50$  Å thick  $\Phi_B = 320$  meV high  $Al_{0.32}Ga_{0.78}As$  tunnel barrier. Both QWs are modulation doped with a Si- $\delta$  layer 500 Å and 400 Å away from the respective QWs. The electron sheet density in the bulk of  $QW^\perp$  and  $QW^{\parallel}$  after illumination is  $2 \times 10^{11} \text{ cm}^{-2}$  for both and they are 5000 Å and 3600 Å below the surface, respectively. The low temperature mobility in  $QW^\perp$  is  $\mu_1 \approx 2 \times 10^6 \text{ cm/Vs}$  while for  $QW^{\parallel}$  we estimate  $\mu_2 \geq 2 \times 10^5 \text{ cm/Vs}$ . In the  $y$ -direction the sample geometry is translationally invariant and the tunnel junction extends about 20 μm in width. The QWs are separately contacted with ohmic indium contacts and the tunnel current  $I$  is studied in a  ${}^3He$  cryostat at temperatures of 400 mK.

## 3 $V = 0$ : structure of a Sharp Edge

Fig. 4 shows the differential tunnel conductance  $\frac{dI}{dV}$  while sweeping the  $B$ -field at several voltage biases  $V$  applied to the probe  $QW^{\parallel}$  with respect to the



**Fig. 2.** (Sample I:  $n^\perp = 1.9 \times 10^{11} \text{ cm}^{-2}$ ) The differential tunnel conductance  $dI/dV$  as a function of the magnetic field  $B$  at DC bias voltages of 0, -12, -22 mV. Negative bias indicates tunneling of electrons into the edge states. Curves shifted by  $3\mu\text{S}$

QH system in  $QW^\perp$ . Negative bias therefore means electron tunneling from the probe contact into the QH edge. At zero bias we observe well developed peaks in  $\frac{dI}{dV}$  at certain values of the magnetic field. We resolve up to 4 of these resonances (denoted with  $n = 0, 1, 2, 3$  from right to left). Their height and width is larger for those at higher  $B$ -field. The peaks show a slightly asymmetric shape with a steeper slope at the high  $B$  side. Above 5 T we observe a strong suppression of the tunnel current at zero bias. With a finite negative DC voltage bias  $V$  the resonances in  $\frac{dI}{dV}$  and the onset of the tunnel suppression shift towards higher magnetic fields.

### 3.1 The Momentum Conservation Picture: $k_y$

In the following we explain the special tunnel selection rules based on momentum conserved tunneling. In the presence of a magnetic field the electronic states in  $QW^\perp$  are quantized to Landau levels with energy gaps proportional to  $B$ . With an applied junction bias  $V$  the confining edge potential  $\Phi(x, V)$  enters into the full Schrödinger equation as

$$\left[ \frac{(\mathbf{p} - e\mathbf{A})^2}{2m^*} + \Phi(x, V) \right] \Psi(x, y) = E\Psi(x, y) . \quad (1)$$

With a Landau gauge  $\mathbf{A}(x, y) = xBy$  where  $x = 0$  in the center of  $QW^\parallel$ , (1) becomes translationally invariant in the  $y$ -direction. Expressing  $\Psi_n(x, y) = \psi_{nk_y}(x)e^{ik_y y}$ , we can solve for the motion in  $x$  as well as the dispersion  $E_n^\perp(k_y)$

$$\left[ \frac{\mathbf{p}_x^2}{2m^*} + \frac{1}{2}m^*\omega_c^2(x - k_y l_B^2)^2 + \Phi(x, V) \right] \psi_{nk_y}(x) = E_n^\perp(k_y)\psi_{nk_y}(x) \quad (2)$$

where  $n$  is the Landau index, and  $l_B = \sqrt{\hbar/eB}$  defines the magnetic length.

For  $QW^{\parallel}$  the  $k$ -space dispersion  $E^{\parallel}(k_y)$  exhibits a parabolic shape with a well defined Fermi point  $FP^{\parallel}$ . From our choice of Landau gauge, the mass parabola will always be centered at  $k_y = 0$ . For this single-particle picture we neglect spin splitting since the *bare* Zeeman energy is below our experimental resolution. Since orbital effects from the in-plane field can be shown to be negligible<sup>1</sup> we neglect the influence of the magnetic field on  $QW^{\parallel}$ .

The translational invariance of the geometry in  $y$ -direction together with the high mobility of the 2DEGs and the high uniformity of the tunnel barrier causes both momentum  $k_y$  and energy  $E$  to be conserved during tunneling. Applying a voltage bias shifts the dispersion curves in energy, and the magnetic field shifts them in momentum space with respect to each other through the Lorentz impulse acquired by tunneling the effective distance  $\bar{x}$  through the barrier:  $\Delta k_y = eB\bar{x}/\hbar$ . Scanning both the voltage bias and the magnetic field we can thus map out the entire  $E^{\perp}$  vs.  $k_y$  space using  $FP^{\parallel}$  as a probe for the Landau level dispersion  $E_n^{\perp}(k_y)$ .

### 3.2 The Orbit-Center Conservation Picture: $X$

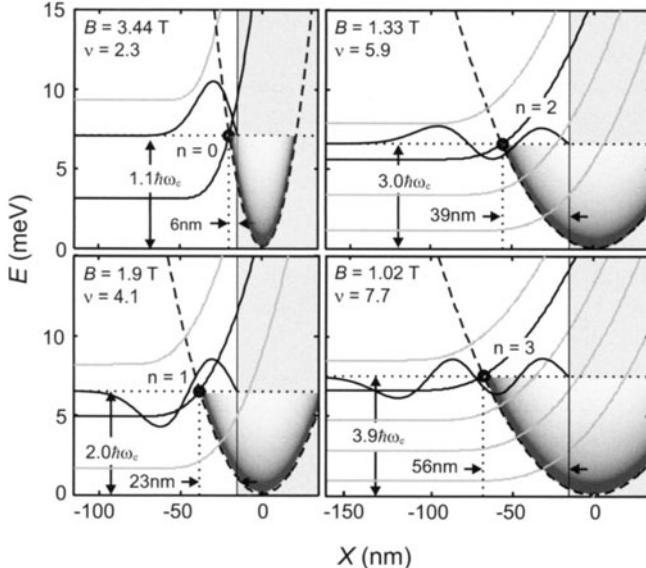
With the orbit center coordinate  $X = l_B^2 k_y$ , which scales with the momentum in  $y$  by the square of the magnetic length,  $l_B^2 = \hbar/eB$ , we can describe the tunnel selection rules also in terms of the conservation of  $X$ .

The wave function can be written as  $\Psi_{n,X}(x, y) = \exp(-iXy/l_b^2)\psi_{n,X}(x)$ , with the  $x$ -component  $\psi_{n,X}(x)$  obeying the Schrödinger equation

$$\left[ \frac{\mathbf{p}_x^2}{2m^*} + \frac{1}{2}m^*\omega_c^2(x - X)^2 + \Phi(x, V) \right] \psi_{n,X}(x) = E_n^{\perp}(X)\psi_{n,X}(x). \quad (3)$$

We can write  $E^{\parallel}$  as a function of  $X$  as well. The dispersion curve  $E_n^{\perp}(X)$  is calculated for an ideally sharp step function edge potential  $\Phi(x)$  at the barrier position, and plotted with  $E^{\parallel}$  in Fig. 3 for four different magnetic fields at  $V = 0$ . Since both energy  $E$  and orbit center  $X$  (or identically transverse momentum  $k_y = X/l_B^2$ ) are conserved upon tunneling, when we plot these two quantum numbers for both systems (Fig. 3) tunneling is only allowed where the two dispersion relations  $E_n^{\perp}(X)$  and  $E^{\parallel}(X)$  intersect.  $dI/dV$  features a maximum whenever one of the Fermi levels coincides with such a crossing point. Therefore we use the Fermi point  $FP$  of the probe  $QW^{\parallel}$  with Fermi momentum  $-k_F^{\parallel}$  as a probe for the spectrum in the QH system in  $QW^{\perp}$ . The appropriate resonance condition is  $E_n^{\perp}(X_F) = E_F^{\perp} - eV$  where the magnetic field tunes the probed position  $X_F = \hbar k_F^{\parallel}/eB$  one cyclotron radius from the center of  $QW^{\parallel}$ , and the tunnel bias  $V$  shifts the dispersion curves in energy

<sup>1</sup> Calculations showed that the  $k$ -space position for  $FP^{\parallel}$  and the dispersion velocity there change less than 5 % even up to 10 T where  $l_B < W$ , so the parabolic dispersion assumption is justified.



**Fig. 3.** Energy diagrams  $E(X)$  of the quantum Hall edge states for the resonance conditions at  $V = 0$  as a function of the orbit guiding center position  $X$ . The black circle indicates the edge state at the Fermi energy that is in resonance with the Fermi point of the probe contact. The wave function of this edge mode is also depicted

with respect to each other. Thus we map out the full dispersion relation  $E_n^\perp(X)$  of the individual edge modes for the  $n$  Landau levels ( $n = 0, 1, 2, \dots$ ) revealed by the maxima in the experimental tunnel conductance data (Fig. 4). Magnetic fields above 4 T separate the occupied states of both systems in the  $X$  space, explaining why tunneling at zero bias is no longer possible.

### 3.3 Real-Space Position of Sharp Edge Channels

In the following we focus on the zero bias conductance and first study the peak positions. By sweeping the  $B$ -field, the *FP* successively scans the current carrying QH edge channels right at the Fermi energy  $E_F^\perp$ . With a known Fermi momentum  $|k_F^\parallel| = 1.1 \times 10^8 \text{ m}^{-1}$  the peak positions directly yield the distance  $X_b = |X_F| - b - \frac{w_\parallel}{2}$  of the edge channel guiding from the tunnel barrier at the respective  $B$ -field.

Table 1 compares this experimental value  $X_b$  with the theoretical sharp edge limit  $X_{b,min}$  determined as follows: When the bulk Fermi energy is pinned within the  $N^{\text{th}}$  bulk Landau level, the position of the  $n^{\text{th}}$  edge is simply given by  $X_{b,min}(N, n) = \xi_{N,n} l_B$ , where  $\xi_{N,m}$  is the  $m^{\text{th}}$  zero of the  $N^{\text{th}}$  order Hermite polynomial and  $n + 1 \leq N$ . Although the two low- $B$  peaks do not strictly satisfy the condition  $n + 1 \leq N$ , disorder induced level-broadening

**Table 1.** Measured distance  $X_b$  of the orbit guiding center from the tunnel barrier for edge channels  $n = 0, 1, 2, 3$  at various  $B$ , compared to the magnetic length  $l_B$  ( $n = 1.9 \times 10^{11} \text{ cm}^{-2}$ )

n	0	1	2	3
$B$ (T)	3.44	1.9	1.33	$\sim 1.02$
$\nu$	2.3	4.1	5.9	7.7
$l_B$ (nm)	14	19	22	25
$X_{b,min}$ (nm)	0	13	27	42
$X_b$ (nm)	$6 \pm 2$	$23 \pm 3$	$39 \pm 4$	$56 \pm 6$

restores the domain of validity of this model. The small difference of the model from the experimental value is consistent with a slight band bending towards the edge of  $QW^\perp$ , which arises from the detailed electrostatics of this non-planar tunnel geometry [14], and the pinning condition of the Fermi energy in the bulk is softened in disorder broadened Landau levels, as already mentioned. Nevertheless the measured distance  $X_b$  is always within  $l_B$  of this sharp edge limit  $X_{b,min}$  and the length scale between edge channels is found to be of order or smaller than both the magnetic length  $l_B$  and the Bohr radius  $a_0 = 10$  nm in GaAs. As a consequence, the formation of compressible strips as predicted by Chklovskii, et al. [15] does not occur here since the electrostatic screening central to the model must take place on length scales larger than  $a_0$ . Integer QH edge reconstruction proposed for soft edges is likewise not expected to occur here since it was shown [16] to occur only for density gradients smoothly varying over a length scale  $8 l_B$  or greater. This quantified observation of the sharp edge limit is one of our main results.

### 3.4 Lineshape Analysis

For the lineshape analysis of the conductance peaks, we note that the tunnel conductance at  $V = 0$  is proportional to the number of states at the Fermi energy that overlap in momentum space,

$$\frac{dI}{dV} \sim |t(B)|^2 \sum_{k_y, k'_y} A_{QHE}(k_y, E_F) A_{probe}(k'_y, E_F) \delta_{k_y, k'_y}. \quad (4)$$

Solving the Schrödinger equation for the transmission we find that the transmission probability  $|t(B)|^2 \sim B$  is roughly proportional to the magnetic field for low fields ( $l_B > W, b$ ), contributing to the lower peak height at lower  $B$ . To quantify the lineshape we assume  $A_{QHE}(k_y, E_F) = \sum_n \delta(E_F - E_n(k_y))$  for the spectral function of the quantum Hall system. In the probe contact the component of the Fermi circle in the  $k'_z$  direction results in a van Hove-like singularity in the probe spectral function at the Fermi point  $k'_y = -k_F$ :  $A_{probe}(k'_y, E_F) = 1/\sqrt{E_F - \hbar^2 k'^2_y/2 m^*}$ . The resulting calculated  $dI/dV$  reproduces the asymmetry observed in the experimental conductance of Fig. 4

(dashed line). As in Ref. [17] we introduce an additional Gaussian broadening to account for small non- $k$ -conservation which shows an excellent phenomenological fit of the experimental data (dash-dotted line).

### 3.5 Substructure beyond Outermost Edge

The additional shoulder at the high  $B$ -field side of the resonance for  $n = 0$  can not be explained within this model. Translated into the orbit-center coordinate  $X$  the shoulder is separated from the main peak by  $\sim 3$  nm. Analogous to the length scale comparisons above, this short distance scale rules out that it is a signature of either standard edge reconstruction [16] or the Chklovskii compressible strips picture.

A possible explanation could be proposed due to spin-splitting of the spin-degenerate Landau band at the edge. Considering exchange interaction Dempsey et al. [18] predicted a spontaneous transition between spin-unpolarized and spin-polarized ground states in the outermost edge state. Within this explanation however we would expect two equally pronounced peaks, one for each spin branch.

### 3.6 Width of Conductance Peak

We now focus on the line width of the conductance peak in magnetic field, to compare with the anomalous results of Kang, et al. [8]. These authors report a zero bias tunnel conductance peak between two quantum Hall edges that extends over an unexpectedly wide filling factor range  $\delta\nu$ . In contrast, in the data of Fig. 2, we see a narrow peak whose linewidth is completely describable with a non-interacting overlap integral of spectral functions, suggesting that in our perpendicular-tunnel geometry, this effects do not occur. The anomalous width was ascribed to a wide tunnel gap  $\Delta$  [11,12,13] where a direct relation  $\delta\nu \approx \Delta/(\hbar\omega_c)$  was assumed. In the following, we remark on a simple non-interacting model which may be able to explain the difference of behavior in the two experiments.

The relation for  $\Delta$  above is based on the assumption of a constant density of states in the bulk, even under quantizing magnetic fields, leaving the Fermi energy constant with  $B$ . However any modulation of the density of states (DOS), for example the onset of formation of bulk Landau levels, would lead instead to oscillations in the Fermi energy with  $B$ . Although the low mobility reported in Kang's samples was interpreted to result in a constant DOS, similarly low mobility 2DEGs were found to show modest oscillations of the chemical potential [14]. Since the chemical potential of the device of Ref. [8] would oscillate in both QH contacts simultaneously, the oscillations would manifest themselves as extended conductance features at magnetic fields wherever the density of states is large, conceivably producing an extended peak width. In our experiment, however, this co-oscillating mechanism is absent, meaning there should be no extended linewidth beyond the single particle picture, as

observed. This feature should be replaced by a structure associated with the jump in the QH chemical potential as the Fermi energy traverses the gaps between filling factors. In fact we find a feature that could be ascribed to this jump, described in detail in Section 4. In summation, this therefore raises the question as to whether the zero bias anomaly in Kang's experiment could just be caused by a Fermi level that happens to be pinned at the resonance condition for a finite range of  $B$ .

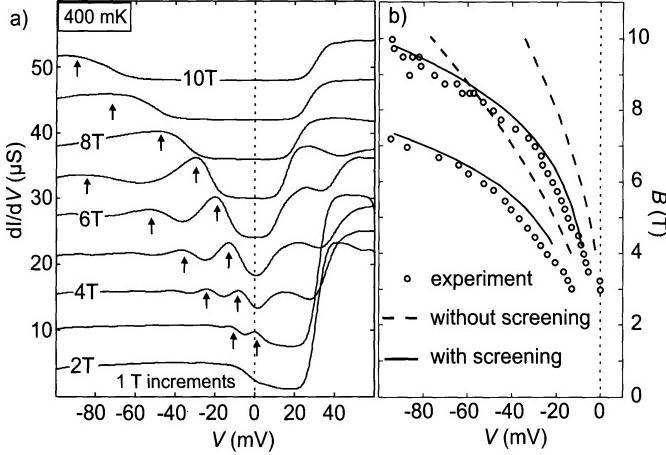
## 4 $V < 0$ : Dispersion of a Sharp Edge

Scanning both the voltage bias and the magnetic field we can map out the entire  $E^\perp$  vs.  $k_y$  space using  $FP^\parallel$  as a probe for the Landau level dispersion  $E_n^\perp(k_y)$ . This dispersion relation at finite tunnel bias  $V$  is the focus in this paragraph. Fig. 4 shows several conductance traces plotted against applied voltage bias for a series of  $B$ -values between 2 T and 10 T. At  $B$  fields above 2.5 T we observe clear maxima and minima at low negative voltages. They become more pronounced and shift towards negative bias at higher magnetic fields, with their separation increasing with magnetic field from about 10 mV at 3.5 T to more than 50 mV at 7 T. As already addressed in paragraph 3 the conductance at zero bias disappears beyond a  $B$ -field of 4 T and the voltage range of suppressed conductance around zero bias increases with higher  $B$ -fields. At 10 T the conductance is suppressed down to almost -60 mV. Magnetic fields above 4 T separate the occupied states of both systems in the  $k_y$  space, explaining why tunneling at zero bias is no longer possible. Spin splitting is not resolved in these measurements.

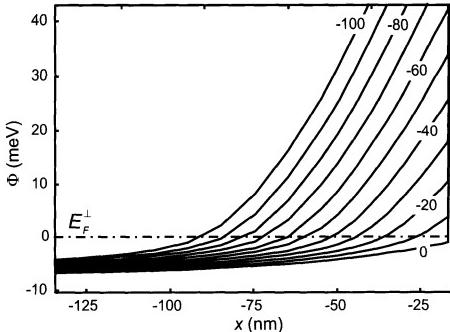
The experimental conductance maxima for tunneling into the lowest two Landau levels are indicated by little arrows in Fig. 4a and their positions are plotted in the  $V - B$  plane (Fig. 4b, circles) for comparison with model calculations based on the equation (2) with a predetermined edge potential  $\Phi(x, V)$ . It can be seen that the step function potential model (dashed line) qualitatively explains the resonance features, though the predicted position is to the right of what is experimentally measured. We can explain this shift with a model that includes screening at the 2D edge.

We perform a self-consistent Schrödinger-Poisson calculation of the edge potential at a biased tunnel junction without magnetic field and plotted the result in Fig. 5. There we have plotted  $\Phi(x, V)$  as the lowest 2D subband in  $QW^\perp$  as a function of position  $x$ . The subband is occupied up to the Fermi energy designated by the dashed line. Even at zero bias the potential close to the edge is not flat. Negative bias lifts the subband energy above the Fermi level resulting in edge depletion of order 750 Å at -100 mV, for example.

Based on the results of the zero  $B$ -field simulation for  $\Phi(x, V)$  we now solve the Schrödinger equation (2) with  $B$  and calculated the position of the expected maxima in the tunnel conductance (Fig. 4b). The self-consistently calculated edge potential shows excellent agreement with experiment.

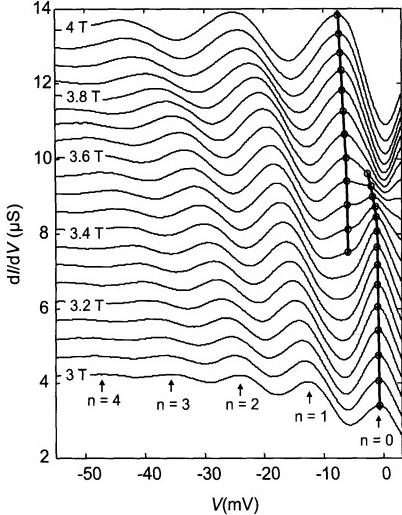


**Fig. 4.** a) Differential conductance  $dI/dV$  plotted against voltage bias ( $V$ ) between  $QW^\perp$  and  $QW^{\parallel}$ , with successive traces shifted by  $+6 \mu\text{S}$ . The resonances due to tunneling into the lowest two Landau levels are marked by arrows. b) Positions of  $dI/dV$  maxima for the lowest two Landau levels plotted in the  $V - B$  plane: The experimental results (*circles*) are qualitatively similar to the values determined from a flat edge potential (*dashed line*) but are much better described using a self-consistently calculated depleted edge potential (*solid line*)



**Fig. 5.** Self-consistently calculated potential  $\Phi(x, V)$  at the edge of  $QW^\perp$  for a series of applied bias voltages  $V$  (denoted in mV) at  $B = 0$ . The potential barrier is located at  $x = -15$  nm. The plotted lines represent the local bottom of the 2D-subband in  $QW^\perp$ . Electronic states are filled up to the Fermi level  $E_F^\perp$ . Note the onset of a depletion region at 10 mV bias

We checked our self-consistent Schrödinger-Poisson calculations against the analytical model of Levitov, et al. [19] which was designed to model a similar device geometry. We find agreement with the zero bias density distribution to within 5% at length scales of 400 Å or more away from the barrier, but at shorter distances and large biases the analytical model appears to underestimate the edge depletion under negative bias as well as the edge accumulation under positive bias. This is presumably a consequence of the finite screening length in  $QW^\perp$ , in contrast with the metallic charge mirror assumed in the analytical model. In this paragraph we discuss an anomaly in the probed dispersion relation of the outermost edge channel near filling factor  $\nu = 2$ . Fig. 6 shows  $dI/dV$  as a function of the tunnel bias  $V$



**Fig. 6.** (Sample II:  $n^\perp = 1.7 \times 10^{11} \text{ cm}^{-2}$ ) The resonance in  $\frac{dI}{dV}(V)$  for the outermost edge channel (Landau index  $n = 0$ ) splits into two peaks near 3.5 T at small negative bias. This is evidence for a jump in the chemical potential expected for a sharp edge near  $\nu = 2$

at a series of fixed  $B$ -fields.  $B \sim 3.5$  T corresponds to bulk filling factor  $\nu = 2$ . Negative  $V$  shows the continuous dispersion for the empty states of the higher Landau levels. However, the dispersion curve of the lowest level does not cross the zero bias line continuously. Instead we observe a splitting  $\Delta V \sim 4$  mV of the peak near  $B \sim 3.5$  T. We note that the splitting  $e\Delta V$  is close to the cyclotron energy  $\hbar\omega_c = 6$  meV at  $B = 3.5$  T. One possible explanation, therefore, is that the bulk goes through the highly incompressible state  $\nu = 2$  the Fermi energy  $E_F^\perp$  transverses the gap  $E_{gap}$  in the density of states between the two lowest Landau levels. Holding up the resonance condition (3.2) requires the readjustment of  $V$  by  $\Delta V = E_{gap}/e$ . Spatial density inhomogeneities support the simultaneous existence of both situations ( $\nu < 2$  and  $\nu > 2$ ) along the tunnel junction at a finite range of the  $B$ -field. The difference  $\hbar\omega_c - E_{gap} \approx 2$  meV is therefore the full width of the broadened Landau levels (including localized and extended states). Thermal activation measurements within  $QW^\perp$  confirmed 6 meV as the transport gap at  $\nu = 2$ .

The absence of such a splitting at  $B = 3.5$  T for the higher Landau levels, i.e. at more negative  $V$ , could be explained with the formation of compressible strips at the depleted edge. The metallic properties of these strips screen the jump of the chemical potential so that it is not observed at the edge.

## 5 Conclusion

In conclusion we have probed the QHE edge state dispersion relation  $E_n^\perp(k)$  at a sharp cleaved edge. In the regime of high magnetic fields and moderate negative bias where only the lowest Landau level is occupied and the probed states are empty. We observe a shift in the location of the conductance maxima that directly reveals the band bending under bias at the tunnel junction,

and is in excellent agreement with a self-consistent model calculation. For the zero bias case we have demonstrated that the Chklovskii picture is not valid here. We have directly measured the real space position of the edge channels. The resonance lineshapes are fully understood if we include a Gaussian broadening in the momentum selection rule and described as results of the spectral functions of the tunnel contacts and there is no additional persistent conductance found at zero bias unlike the observations by Kang et al. [8] in a coplanar double edge tunneling experiment. This proves that such a cleaved quantum Hall edge allows an effective filling factor at the edge to be the same as that in the quantum Hall bulk. This is an important condition for the interpretation of a series of previous tunnel experiments on cleaved quantum Hall edges [1,2,3,4,20]. Zuelicke, et al. [10] predict that in the tunneling geometry used in the present experiment, information about the 2D spectral function in fractional QH edges can be learned.

### Acknowledgements

We gratefully thank U. Zülicke for stimulating discussions. The work was supported by DFG via Schwerpunktprogramm "Quanten-Hall-Systeme".

## References

1. A. Chang, L. N. Pfeiffer, and K. W. West, Phys. Rev. Lett. **77**, 2538 (1996).
2. M. Grayson, D. C. Tsui, L. N. Pfeiffer, K. W. West, and A. M. Chang, Phys. Rev. Lett. **80**, 1062 (1998).
3. A. M. Chang, C. C. C. M. K. Wu, L. N. Pfeiffer, and K. W. West, Phys. Rev. Lett. **86**, 143 (2001).
4. M. Hilke, D. C. Tsui, M. Grayson, L. N. Pfeiffer, and K. W. West, Phys. Rev. Lett. **87**, 186806 (2001).
5. S. Q. Murphy, J. P. Eisenstein, L. N. Pfeiffer, and K. W. West, Phys. Rev. B **52**, 14825 (1995).
6. B. Kardynal, C. H. W. Barnes, E. H. Linfield, D. A. Ritchie, J. T. Nicholls, K. M. Brown, G. A. C. Jones, and M. Pepper, Phys. Rev. B **55**, R1966 (1997).
7. O. M. Auslaender, A. Yacoby, R. de Picciotto, K. W. Baldwin, L. N. Pfeiffer, and K. W. West, science **295**, 825 (2002).
8. W. Kang, H. L. Stormer, L. N. Pfeiffer, K. W. Baldwin, and K. W. West, nature **403**, 59 (2000).
9. A. Altland, C. H. W. Barnes, F. W. J. Hekking, and A. J. Schofield, Phys. Rev. Lett. **83**, 1203 (1999).
10. U. Zülicke, E. Shimshono, and M. Governale, Phys. Rev. B **65**, 241315 (R) (2002).
11. A. Mitra and S. M. Girvin, Phys. Rev. B **64**, 041309(R) (2001).
12. S. Nonoyama and G. Kirczenow, Phys. Rev. B **66**, 155334 (2002).
13. M. Kollar and S. Sachdev, Phys. Rev. B **65**, 121304 (2002).
14. M. Huber, M. Grayson, M. Rother, R. A. Deutschmann, W. Biberacher, W. Wegscheider, M. Bichler, and G. Abstreiter, Physica E **12**, 125 (2002).

15. D. B. Chklovskii, B. I. Shklovskii, and L. I. Glazman, Phys. Rev. B **46**, 4026 (1992).
16. C. de C. Chamon and X. G. Wen, Phys. Rev. B **49**, 8227 (1994).
17. J. Eisenstein, T. J. Gramila, L. N. Pfeiffer, and K. W. West, Phys. Rev. B **44**, 6511 (1991).
18. J. Dempsey, B. Y. Gelfand, and B. I. Halperin, Phys. Rev. Lett. **70**, 3639 (1993).
19. L. S. Levitov, A. V. Shytov, and B. I. Halperin, Phys. Rev. B **64**, 075322 (2001).
20. M. Grayson, D. C. Tsui, L. N. Pfeiffer, K. W. West, and A. M. Chang, Phys. Rev. Lett. **86**, 2645 (2001).

# Structural Properties of Semiconductor Nanostructures from X-Ray Scattering

J. Stangl<sup>1</sup>, T. Schülli<sup>1</sup>, A. Hesse<sup>1</sup>, V. Holy<sup>2</sup>, G. Bauer<sup>1</sup>,  
M. Stoffel<sup>3</sup>, and O.G. Schmidt<sup>3</sup>

<sup>1</sup> Johannes Kepler Universität  
4040 Linz, Austria

<sup>2</sup> Masaryk University  
601 77 Brno, Czech Republic

<sup>3</sup> Max Planck-Institut für Festkörperforschung  
70569 Stuttgart, Germany

**Abstract.** The optical and electrical properties of self-assembled semiconductor nanostructures are strongly affected by their chemical composition and strain state. These depend sensitively on the growth conditions. X-ray diffraction techniques are used in order to investigate series of SiGe/Si(001) island samples grown by molecular beam epitaxy at various temperatures. The chemical composition distribution and the in-plane strain of uncapped islands are obtained from x-ray diffraction reciprocal space maps and from iso-strain scattering combined with anomalous scattering at the Ge K-edge. From reciprocal space maps around asymmetric Bragg reflections, also the properties of capped islands are obtained, giving insight into the structural changes during capping of SiGe islands with Si.

## 1 Introduction

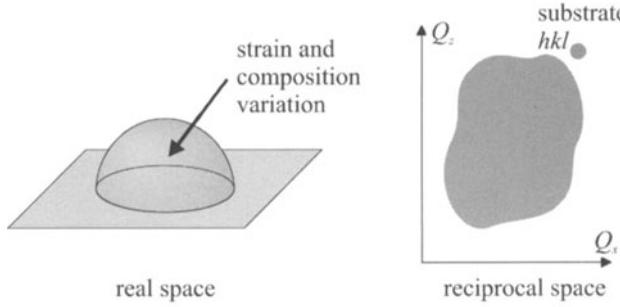
Recently, self-assembled semiconductor islands have attracted attention due to their band-gap modulation, leading to carrier confinement in spatial regions comparable to their de Broglie wavelength in all three dimensions. This opens the possibility of band-gap engineering beyond those in bulk semiconductors and two-dimensional thin films. During growth by molecular beam epitaxy (MBE) or chemical vapor depositions (CVD), various nanostructure shapes have been observed depending on the particular growth conditions [1,2,3,4]. Also the chemical composition of such self-assembled structures may differ considerably from that of the deposited material, again with a strong influence of growth conditions such as substrate temperature and adatom flux. It is hence of importance to correlate the growth conditions with the structural parameters (shape, size, strain state, chemical composition) and the electronic and optical properties of such nanostructures.

X-ray scattering techniques have proved to be powerful for the investigation of nano-scale islands. Using grazing incidence small angle x-ray scattering, the shape and positional correlation of small islands on a surface, or embedded into a superlattice, have been determined [5,6,7]. From x-ray diffraction (XRD), the average chemical composition in such islands has been

established, and using grazing incidence diffraction (GID) also the in-plane relaxation has been determined [8]. Using the iso-strain scattering technique (see below), the distribution of In in InGaAs islands on GaAs(001) was determined with a height resolution of about 1 nm [9,10], the method was also applied to SiGe islands on Si(001). Similar results have been obtained in studies using reciprocal space mapping around asymmetrical Bragg reflections and finite element simulations based on models of nano-scale islands [11,12,13]. Recently, the use of anomalous scattering further advanced these studies [14,15]. Here, we want to present results obtained with both methods on the same SiGe islands grown by MBE on Si(001). The influence of growth temperature on intermixing during growth is addressed, and, as for applications only nanostructures embedded in a matrix can be employed, also the influence of capping at different growth conditions on the structural properties of SiGe islands has been investigated. It can be expected that the interdiffusion during capping is more pronounced than during deposition of Ge, as Ge shows segregation in Si, which counteracts the intermixing during Ge growth, but enhances intermixing during capping.

## 2 Coplanar X-Ray Diffraction

To tackle the issue of island composition we measure reciprocal space maps (RSMs) around asymmetrical reflections in coplanar XRD geometry. In this geometry, the incident beam, the diffracted beam, and the surface normal lie in one plane. The momentum transfer  $\mathbf{Q} = (Q_x, 0, Q_z)$  has an in-plane component as well as a vertical component. Hence the lattice parameters parallel to the sample surface *and* in growth direction are measured, allowing for a reconstruction of the unit cell, and hence the strain state and the composition of the unit cells in the island. Each "unit cell" produces a scattered intensity distribution spread in reciprocal space, and the contributions of all unit cells sum up coherently. In this geometry, a simple separation of the form factors of the individual strain states is not possible, as sketched in Fig. 1. Instead, simulations are performed starting from model assumptions of the whole island: Based on a certain Ge distribution in the island and assumptions on the island shape, finite element method-based calculations are used to determine the three-dimensional strain state of the island. Details on these calculations can be found in Ref. [13]. From the strain field, the scattering pattern is calculated. The result is compared to the experimentally obtained diffuse intensity distribution, and the model is adjusted until a good correspondence is found. The obvious disadvantage of this method is a high computational effort, and a rather indirect way of obtaining the sample parameters. An advantage of the method is, however, that it is not restricted to islands on a flat surface, but allows also the analysis of buried structures, which is so far not possible using iso-strain scattering (see below).



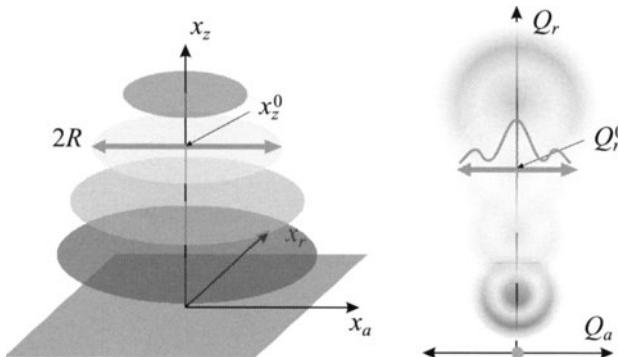
**Fig. 1.** Sketch of an interdiffused island and corresponding intensity distribution in coplanar XRD

For the simulations, we assume cylindrically symmetric islands with the shape of a rotational paraboloid. For uncapped islands, the base radius  $R$  and the height  $H$  are used as obtained from AFM measurements. The Ge content is assumed to vary vertically from a value  $x_1$  at the island bottom to  $x_2$  at the island apex; for its variation we assume a square-root behavior, i.e., the content is increasing faster at the base than at the top of the islands:  $x_{\text{Ge}}(z) = x_1 + (x_2 - x_1) \cdot \sqrt{z/H}$ . With this assumption, the strain field is calculated using the finite element method (FEM) for a 45° wedge through the island and the substrate (and cap layer for capped islands), extending from the [110] to the [100] in-plane direction. The mesh used for the calculation is very fine within and close to the islands, where large strain values occur, and continuously less dense with increasing distance from the island, where strain fields are weak. The outer limit of the mesh, where the nodes cannot move perpendicular to the boundaries, are chosen so that a further extension of the calculated range does not change the obtained strain fields.

From the strain field obtained by FEM, the diffraction pattern are calculated using the distorted wave Born approximation, and  $x_1$  and  $x_2$  are then varied until a good correspondence between measurement and simulation is achieved (for more details see Ref. [13] and references therein). For capped islands, the values for  $R$  and  $H$  for uncapped islands grown under the same growth conditions are used as a starting point. Then these values are changed as well in order to obtain a good correspondence between simulation and experiment.

### 3 Iso-Strain Scattering

An alternative approach is based on grazing incidence diffraction (GID). Here, the primary beam impinges on the sample surface under a very shallow angle  $\alpha_i$  in the range of the critical angle of external reflection  $\alpha_c$ , which is of the order of 0.2° for SiGe and x-ray energies in the range of 8 to 12 keV, as used here. The beam is diffracted at lattice planes perpendicular to the sample surface, and simultaneously reflected at the surface, so that the exit angle to the surface  $\alpha_f$  is again comparable to  $\alpha_c$ . Therefore the momentum



**Fig. 2.** Sketch of the iso-strain areas in an uncapped island and the corresponding intensity distribution in reciprocal space

transfer  $\mathbf{Q}$  is mainly within the growth plane, whereas the component along growth direction  $Q_z$  is very small. Consequently, this scattering geometry is sensitive only to the in-plane component of the lattice parameter, i.e., to in-plane strain.

### 3.1 Interdependence of Strain, Lateral Size, and Height

The iso-strain scattering (ISS) technique was developed by Kegel *shape et al.* to analyze the interdependence of strain, lateral width and height of nano-scale islands [9,10,16]. The basic idea is a decomposition of the scattered signal from a gradually relaxed structure into contributions from areas with constant strain, as sketched in Fig. 2. The *shape in-plane* lattice parameter (only this is accessible in GID) varies monotonically from the substrate value at the island bottom to that of a more or less relaxed material at the top of the island. Each height in the island has a certain lattice parameter, and scattering from that region will appear at a certain value of  $Q_r$  in reciprocal space, where  $Q_r$  is the so-called radial direction from the origin of reciprocal space to the Bragg point under investigation. In a scan along  $Q_r$  the length of  $\mathbf{Q}$  varies, and hence different positions correspond to different in-plane lattice parameters. Each individual iso-strain area gives rise to a certain form factor of the scattered intensity distribution, which overlaps in reciprocal space with the scattering from other iso-strain areas, so that a continuous intensity distribution between the reciprocal lattice point of the substrate  $Q_{r,\text{sub}}$  and that of the relaxed material at the island top  $Q_{r,\text{top}}$  is formed, as shown in Fig. 2.

The lateral extent of a certain iso-strain area can be obtained from scans along  $Q_a$  at the respective value of  $Q_r$ .  $Q_a$  is the so-called angular direction, perpendicular to  $Q_r$ , i.e., in a scan along  $Q_a$  the length of  $\mathbf{Q}$  is constant, and the lateral width of an iso-strain area can be measured from the width and of the central maximum and the distance of form factor oscillations in the  $Q_a$ -scan. Note that the position  $Q_r$  corresponds to an in-plane lattice parameter  $a_{\parallel}$ ; at this point we cannot disentangle the change of in-plane lattice

parameter due to elastic relaxation and the change due to interdiffusion of Si and Ge.

In order to measure the height of an iso-strain area above the substrate surface, we compared the lateral extent of the area to the island profile obtained by AFM. A disadvantage of this method is that due to surface oxidation the shape obtained by AFM is usually slightly larger than the "crystalline island core" observed in x-ray diffraction, so that the obtained heights are slightly underestimated.

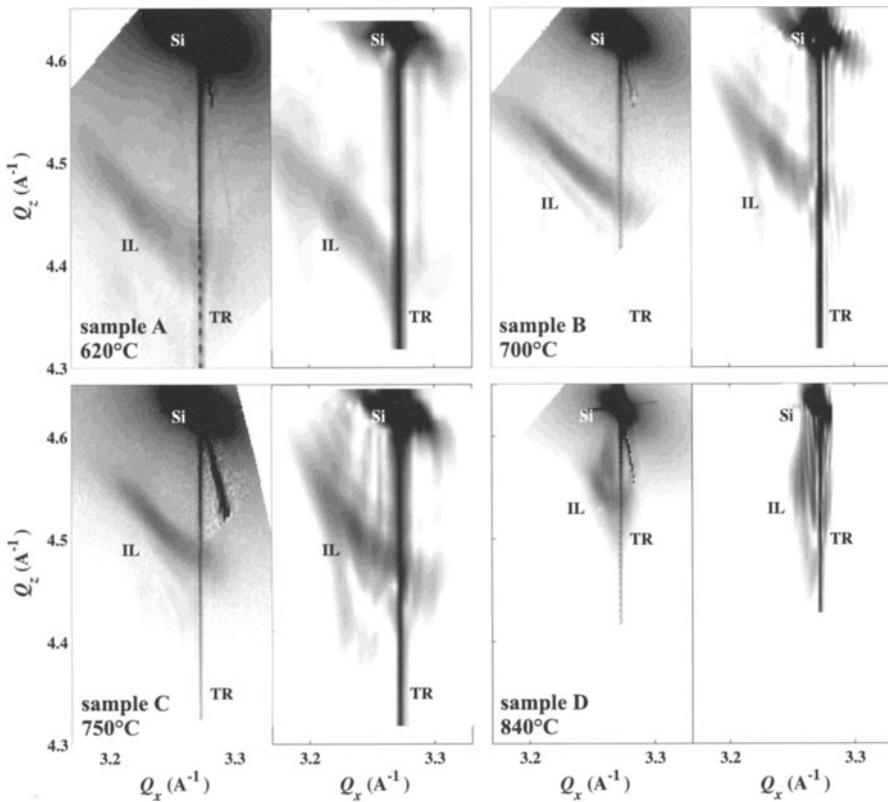
### 3.2 Composition from Anomalous Scattering

So far, from ISS the lateral dimension and height within the island of a region with a certain lattice parameter can be established. To obtain information also on the composition, anomalous scattering has been employed. The atomic scattering factor  $f(Q, E)$  for x-rays is modulated for x-ray energies in the vicinity of an atomic energy level [15]. For SiGe, we vary the energy in the vicinity of the Ge K-edge at 11.104 keV. To determine the composition of a SiGe alloy, for known variation of  $f(Q, E)$  for Ge with energy (as Si has no absorption edge in the energy range around 11 keV, the scattering factor of Si is independent of energy) the scattered signal is measured at two energies, about 1 eV below the edge and about 60 eV below the edge. The change in the scattered intensity for the two energies can then be attributed to a certain Ge concentration [15]. The measurement is performed for the range of  $Q_r$  as in ISS, which links the lattice parameter to a certain Ge content. With this method, the Ge distribution within the islands is established.

## 4 Experiment

We investigated two sets of samples grown by MBE on Si(001) substrates: in the first set (samples A to D), 6 to 11 monolayers (ML) of Ge were deposited onto Si for various growth temperatures from 620°C to 840°C. From AFM investigations of the surface after growth, the formation of dome-shaped islands is observed, with an island size increasing with growth temperature. This indicates that the intermixing of Ge and Si in the islands gets more pronounced with increasing temperature, which is also supported by the fact that for higher growth temperature more Ge needs to be deposited until islands form, indicating a reduced strain of the wetting layer due to intermixing. Reciprocal space mapping and ISS have been employed in order to quantify the Ge content within the islands.

The second set of samples (E to G) consist of SiGe islands formed by the deposition of 5 ML of Ge at 630°C. The islands were capped with 150 nm of Si at substrate temperatures between 450°C and 630°C. Reciprocal space mapping has been used in order to determine the different interdiffusion between Si and Ge during capping with Si. The growth parameters are listed in Tab. 1.

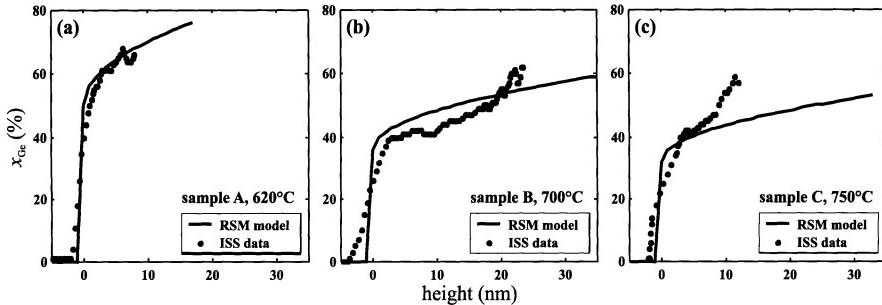


**Fig. 3.** Measured RSMs around the (224) Bragg reflections of the investigated samples (left) together with simulations based on the parameters given in Tab. 1 (right panels)

#### 4.1 Influence of Ge Deposition Temperature

For the first sample series, RSM and ISS measurements have been performed at ESRF in Grenoble, France. For the second set of samples, RSMs around the (224) Bragg reflection have been recorded at the ESRF, as well as at Hasylab in Hamburg, Germany.

Figure 3 shows the measured reciprocal space maps of the samples of the first series (left panels), together with simulations (right panels). Beside the substrate reflection labeled "Si" and the truncation rod ("TR"), the diffuse intensity distribution due to the partially elastically relaxed SiGe islands ("IL") is visible. Qualitatively it can be seen that for the sample grown at higher temperature, the peak due to the islands is closer to that of the Si substrate, indicating a stronger interdiffusion in the islands. With the method described above, we have determined the Ge content profiles within the values  $x_1$  and  $x_2$  as shown in Tab. 1.



**Fig. 4.** Ge composition as a function of height in the island as obtained from RSMs and ISS for various deposition temperatures

The results are corroborated by the Ge profiles obtained from ISS and anomalous scattering. A comparison of the Ge profiles obtained from RSMs and ISS is shown in Fig. 4. For this comparison, the height scales of the two methods have been adjusted (shifted) by few nm. This may be done, as in both methods the actual origin of height has some uncertainty: in the models used for the simulations of RSMs, a sharp interface to the wetting layer with the same composition as the base of the island,  $x_1$ , is assumed, which is certainly a simplification. Actually rather a steep, but still gradual transition from pure Si to SiGe has to be expected. While ISS directly obtains the Ge content without a model, and thus reproduces the gradual increase of Ge content, the height information is obtained from a comparison with AFM data, with an inherent uncertainty as explained above. Both methods assume laterally smooth layers and neglect lateral variations in the Ge composition, which is certainly a simplification as well. A comparison with studies of the lateral profiles shows, however, that these variations are of the order of our error bars (typically 5–10% absolute) for the major part of the islands [17]. For the lower growth temperatures, a very good agreement of both methods can be observed. For higher temperatures, due to the reduced strain in the islands, the error bars of both methods increase, and the quantitative correspondence gets somewhat worse, although the overall results are still in good agreement.

Obviously, an increase in growth temperature leads to a strong enhancement of Si–Ge intermixing, with the maximum Ge content of about 75% at 620° growth temperature dropping to about 40% at 840°C. For the highest deposition temperature (sample D), a qualitative difference compared to the other samples is observed: The aspect ratio  $H/R$  is considerably lower than for the other samples of the series, and the reciprocal space map shows a peak very close to that of the Si truncation rod, i.e., the elastic relaxation is significantly smaller than even for sample C grown at 750°C. The shape obtained from AFM resembles rather a truncated pyramid than a dome with steep side facets. We believe that this change is directly related to a change of the growth mode. As was shown by Jesson *shape et al.*, for deposition of Ge on Si at low growth temperatures, island grow by a statistical nucleation

process [18], i.e., only islands that exceed a critical nucleus size continue to grow, while smaller nuclei dissolve. In contrast, for the deposition of alloys with small strain (small Ge content in our case), a barrierless island formation was predicted [19,20,21]. Although we cannot access the growth kinetics in our ex-situ experiment, the results could be consistent with a change in growth mode due to intermixing. Increasing the growth temperature might have a similar effect than reducing the Ge content of the deposited alloy: due to intermixing, the material forming the island is actually a SiGe alloy, and for 840°C the strain might be low enough to allow for the barrier-less formation of islands.

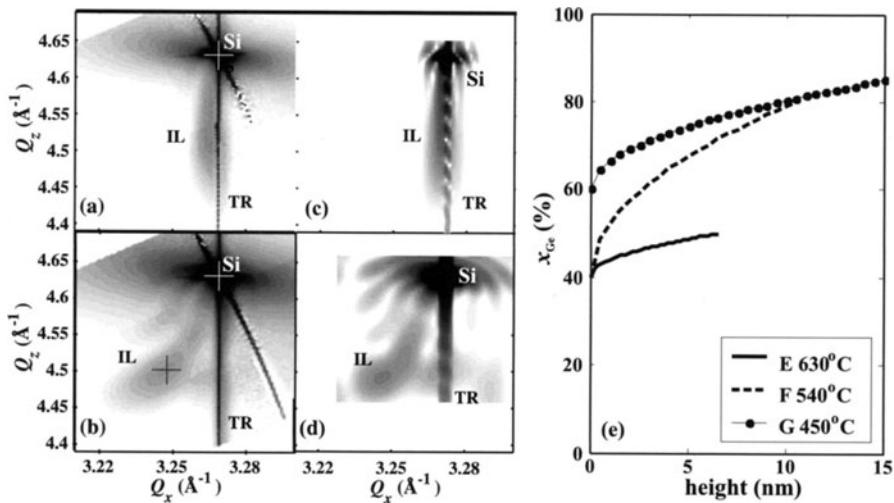
**Table 1.** Growth parameters of the investigated samples and experimental results obtained from reciprocal space maps and iso-strain scattering

Sample	$T_{\text{growth}}$ (°C)	$T_{\text{cap}}$ (°C)	$R_{\text{base}}$ (nm)	$H$ (nm)	$x_1$ (%)	$x_2$ (%)	$x_{\text{max,ISS}}\text{ (%)}$
A	620	–	37	19	50	76	73
B	700	–	50	35	36	59	61
C	750	–	70	33	32	53	57
D	840	–	180	30	27	41	27
E	630	630	50	6.6	40	50	–
F	630	540	42	10.5	40	80	–
G	630	450	35	15	60	85	–

## 4.2 Influence of Capping Temperature

Reciprocal space mapping has also been applied in order to investigate the change in island composition during capping of SiGe islands with Si. For samples E through G, the deposition parameters of Ge are very similar to those of sample A. After the Ge deposition, the islands have been capped with Si at different temperatures. Figure 5 shows the reciprocal space maps together with simulation of samples E and G, capped at 630°C and 450°C, respectively. For sample E, the map shows a diffuse peak due to the islands very close to the truncation rod. Hence almost no elastic relaxation is present in this sample. From the peak shape, narrow along  $Q_x$  but elongated along  $Q_z$ , it can be seen that the islands have a rather flat shape with a large base width (see also the results of the simulations in Tab. 1). Hence during capping at a temperature where a comparably moderate intermixing is observed during Ge deposition, maintaining a high Ge content at the island apex, a significant dilution takes place.

In order to obtain a high Ge content in buried islands after capping with Si, the temperature has to be considerably lowered during capping. For sam-



**Fig. 5.** Measured RSMs of samples E (a) and G (b) together with simulations (panels b,d). The obtained Ge profile for samples E,F, and G is shown in (e)

ple G capped at 450°C, comparable structural properties are observed as for the uncapped sample A grown at 630°C (see Tab. 1 and Fig. 5e). That the Ge content observed in sample G is even slightly higher than for sample A has to be attributed to slightly different growth conditions of both samples, as the two sample series have not been grown in one row. Within one series, we always observed intermixing during capping, but never a Ge enrichment.

#### 4.3 Strain Distribution

The strain distribution in all islands can be directly obtained from the FEM calculations of the island models. They are summarized in Tab. 2. Note that here we give strain values with respect to the Si substrate lattice parameter, i.e.,  $\varepsilon_{xx,zz} = (a_{\parallel,\perp} - a_{\text{Si}})/a_{\text{Si}}$ .

**Table 2.** Strain values for the investigated sample series of uncapped islands

Sample	$\varepsilon_{xx,\text{max}}$ (%)	$\varepsilon_{xx,\text{min}}$ (%)	$\varepsilon_{zz,\text{max}}$ (%)	$\varepsilon_{zz,\text{min}}$ (%)
A	2.1	-2.7	3.8	-0.6
B	1.8	-2.7	3.2	-0.5
C	1.7	-2.7	3.0	-0.5
D	1.0	-1.0	2.2	-0.1

Also from this table, the qualitative difference of sample D is obvious. For the capped samples F through G, the strain values are significantly different due to the fact that the Si cap imposes a strain also from the top. Therefore, the elastic relaxation of uncapped islands is significantly reduced, and the remaining strain depends sensitively on the island shape: for capping at 630°C, the islands flatten to less than half of their height, leading to a reduction of the maximum in-plane strain to a value as low as  $\varepsilon_{xx} \simeq 0.06\%$ . Capping at 460°C almost maintains the aspect ratio of uncapped islands, and the resulting maximum in-plane strain at the island apex is about 0.5%. Although this value is much lower than in uncapped islands, it is still of relevance for device applications, as the Si above the island is strained to the same in-plane lattice parameter. This tensile strained Si region may be used to localize electrons in the Si conduction band, and the strain value is relatively high compared to alternative concepts to achieve strained Si.

## 5 Summary

We have investigated the influence of the growth temperature on the interdiffusion of Si and Ge during SiGe island growth and island capping. While for deposition of Ge onto Si(001) at 620°C the Ge content reaches about 80% in the island apex, growth at higher temperatures leads to strong intermixing. At 840°C, the maximum concentration of Ge in the island apex is only about 30%. During capping of Ge-rich islands grown at 630°C, a further interdiffusion takes place. Capping at 630°C leads to a decrease of the maximum Ge content to about 50%, while lowering the temperature to 450°C during capping maintains the high Ge composition.

## Acknowledgements

This project has been supported by the European Commission (HPRN-CT-1999-00123, HPRI-CT-1999-00040/2001-00140) and FWF (14684). We thank the staff of beamlines ID01 and ID10 (ESRF Grenoble), and W1 (HASYLAB Hamburg) for assistance in the measurements.

## References

1. R. Nötzel, *Semicond. Sci. Technol.* **11**, 1365 (1996).
2. P. M. Petroff, G. Medeiros-Ribeiro, in: *MRS Bulletin: Strain and Heteroepitaxy: Applications to Electronic Materials*, L. Schowalter (Ed.) (Mat. Res. Soc. 1996) p. 50.
3. K. Brunner, *Rep. Prog. Phys.* **65**, 27 (2002).
4. V. Shchukin, N. N. Ledentsov, D. Bimberg, *Epitaxy of Nanostructures* (Springer, Berlin, Heidelberg 2003).
5. M. Schmidbauer, Th. Wiebach, H. Raidt, M. Hanke, R. Köhler, H. Wawra, *Phys. Rev. B* **58**, 10523-10531 (1998).

6. M. Rauscher, R. Paniago, H. Metzger, Z. Kovats, J. Domke, J. Peisl, H.D. Pfannes, J. Schulze, I. Eisele, *J. Appl. Phys.* **86**, 6763 (1999).
7. J. Stangl, V. Holý, P. Mikulik, G. Bauer, I. Kegel, T.H. Metzger, O.G. Schmidt, C. Lange, K. Eberl, *Appl. Phys. Lett.* **74**, 3785 (1999).
8. A.A. Williams, J.M.C. Thornton, J.E. Macdonald, R.G. van Silfhout, J.F. van der Veen, M.S. Finney, A.D. Johnson, C. Norris, *Phys. Rev. B* **43**, 5001 (1991).
9. I. Kegel, T.H. Metzger, A. Lorke, J. Peisl, J. Stangl, G. Bauer, J.M. Garcia, P. M. Petroff, *Phys. Rev. Lett.* **85**, 1694 (2000).
10. I. Kegel, T.H. Metzger, A. Lorke, J. Peisl, J. Stangl, G. Bauer, K. Nordlund, W.V. Schoenfeld, P. M. Petroff, *Phys. Rev. B* **84**, 035318 (2001).
11. Th. Wiebach, M. Schmidbauer, M. Hanke, H. Raidt, R. Köhler, H. Wawra, *Phys. Rev. B* **61**, 5571 (2000).
12. J. Stangl and A. Daniel, V. Holý, T. Roch, G. Bauer, I. Kegel, T.H. Metzger, Th. Wiebach, O.G. Schmidt, K. Eberl, *Appl. Phys. Lett.* **79**, 1474 (2001).
13. A. Hesse, J. Stangl, V. Holý, T. Roch, G. Bauer, O.G. Schmidt, U. Denker, B. Struth, *Phys. Rev. B* **66**, 085321 (2002).
14. R. Magalhaes-Paniago, G. Medeiros-Ribeiro, A. Malachias, S. Kycia, T. I. Kamins, R. Stanley Williams, *Phys. Rev. B* **66**, 245312 (2002).
15. T.U. Schülli, J. Stangl, Z. Zhong, R. T. Lechner, M. Sztucki, T.H. Metzger, G. Bauer, *Phys. Rev. Lett.* **90**, 066105 (2003).
16. I. Kegel, T.H. Metzger, P. Fratzl, J. Peisl, A. Lorke, J.M. Garcia, P.M. Petroff, *Europhysics Lett.* **45**, 222 (1999).
17. A. Malachias, S. Kycia, G. Medeiros-Ribeiro, R. Magalhaes-Paniago, T.I. Kamins, R. Stanley Williams, *Phys. Rev. Lett.* **91**, 176101 (2003).
18. D.E. Jesson, M. Kaestner, B. Voigtländer, *Phys. Rev. Lett.* **84**, 330 (2000).
19. P. Sutter, M.G. Lagally, *Phys. Rev. Lett.* **84**, 4637 (2000).
20. J. Tersoff, B.J. Spencer, A. Rastelli, H. von Känel, *Phys. Rev. Lett.* **89**, 196104 (2002).
21. R.M. Tromp, F.M. Ross, M.C. Reuter, *Phys. Rev. Lett.* **84**, 4641 (2000).

## Part V

### **Quantum Correlations beyond the Fermi Liquid**

# Spectroscopic Signatures of Spin-Charge Separation in a Quasi-One-Dimensional Organic Conductor

Ralph Claessen and Michael Sing

Institut für Physik, Universität Augsburg  
86135 Augsburg, Germany

**Abstract.** In interacting one-dimensional (1D) metals conventional Fermi liquid behavior is expected to break down due to a dynamical decoupling of charge and spin degrees of freedom. Angle-resolved photoelectron spectroscopy (ARPES) on the electronic structure of the quasi-1D organic conductor TTF-TCNQ indeed reveals significant discrepancies to band theory. Instead, the experimental spectra can be well explained by the 1D Hubbard model, which accounts for the intramolecular Coulomb interaction and predicts signatures of spin-charge separation over the entire conduction band width. The model description can even be made quantitative, if one accounts for an enhanced hopping integral at the surface related to a relaxation of the topmost molecular layer. The importance of strong 1D correlation effects is further supported by a remarkable temperature dependence of the ARPES data. These results provide strong spectroscopic evidence for the occurrence of spin-charge separation in the finite energy physics of TTF-TCNQ. Deviations of the Hubbard model description for the *low-energy* spectral behavior is attributed to the neglect of interchain coupling, long-range Coulomb interaction and/or electron-phonon coupling.

## 1 Introduction

In conventional metals the conduction electrons seem to be essentially independent of each other, although as charged particles they should experience a strong Coulomb repulsion. This remarkable observation can be understood on the grounds of the famous Fermi liquid concept in which the low-lying excitations of the many-electron system are described as non or only weakly interacting *quasiparticles*. It can be shown that this picture breaks down, if the electrons are confined to only one spatial dimension. For one-dimensional (1D) metals the electronic charge and spin degrees of freedom are found to decouple from each other, leading to a new low-energy phenomenology referred to as Tomonaga-Luttinger liquid (TLL) [1]. While the theoretical understanding of TLL behavior is well-established by now, direct experimental evidence is still very scarce. Furthermore, so far the search for unusual properties in (quasi-)1D metals has been directed at their low-energy properties where other effects may interfere (*e.g.*, electron-phonon coupling). Recent theoretical treatments of the 1D Hubbard model [2,3] however suggest that

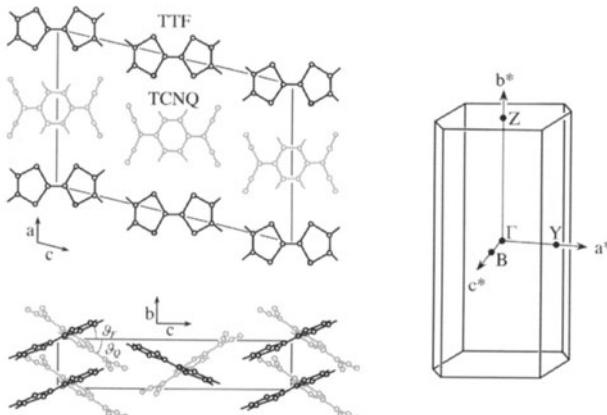
signatures of spin-charge separation may occur even at larger energies comparable with the conduction band width. Here we report on results from angle-resolved photoelectron spectroscopy (ARPES) on the quasi-1D organic conductor TTF-TCNQ [4,5] which are in excellent agreement with this theoretical picture.

## 2 Organic Charge-Transfer Salts as Strongly Correlated Quasi-1D Metals

Nature offers us a number of strongly anisotropic materials whose electronic properties can be viewed as quasi-1D. Among those the so-called organic charge-transfer salts are particularly interesting. Their basic building blocks are organic molecules of planar geometry. Due to the  $\pi$ -symmetry of their highest occupied/lowest unoccupied molecular orbital (HOMO/LUMO), covalent bonding is only possible perpendicular to the molecular plane, which results in the formation of molecular stacks (see Fig. 1). Thus the electrons can propagate only along the stack direction. Often the HOMO of the isolated molecule is fully occupied, so that charge transfer to (or from) suitable acceptor (or donor) complexes is required to render the stacks metallic. In the case of TTF-TCNQ, whose structure is shown in Fig. 1, there are actually two types of molecules, TTF and TCNQ, which form separate and parallel stacks. In this case TTF acts as donor and TCNQ as acceptor, with a charge transfer of 0.59 electrons per formula unit from TTF to TCNQ [6]. Note also, that the molecules are not oriented strictly perpendicular to the stack axis (which is the crystallographic b-direction) but are slightly tilted.

The electronic properties of TTF-TCNQ show indeed a strong quasi-1D character. For example, for pure samples the conductivity along the b-axis is more than 1000 times larger than in any other direction. Below 54 K TTF-TCNQ is a Peierls insulator with a  $2k_F$  charge density wave (CDW). The metallic phase at higher temperatures deviates in many respects from simple band behavior [6]. For example, the magnetic susceptibility is strongly enhanced compared to a band-like Pauli paramagnetism. A dominating role of electronic correlation effects has also been inferred from the observation of  $4k_F$  fluctuations in diffuse x-ray scattering. This suggests that a 1D Hubbard model may provide a good description of the electronic structure. Here the model parameters  $t$  and  $U$  describe the hopping between neighboring molecules along the stacks and the intramolecular Coulomb energy, respectively.

Recently, optical spectroscopy [7] and transport measurements [8] on the closely related Bechgaard salts have given some indirect evidence of possible TLL behavior. The most direct signatures for spin-charge separation are expected to show up in the momentum-resolved single-particle excitation spectrum  $A(k, \omega)$  [1,2,3], which can in principle be measured by ARPES. However, as the photoelectrons have a rather low probing depth ( $\sim 10$  Å), this method

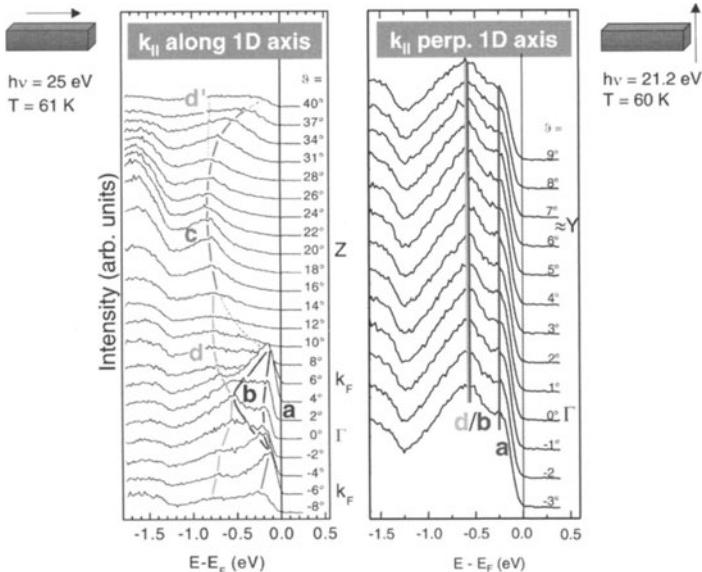


**Fig. 1.** Monoclinic crystal structure and Brillouin zone (with high-symmetry points) of TTF-TCNQ

requires a good control of quality and long-term stability of the surfaces on a molecular level. This has proven to be problematic for the Bechgaard salts due to the polarity of their surfaces and the rapid radiation damage in the exciting light field [9]. In contrast, the surfaces of TTF-TCNQ crystals are non-polar and hence unreconstructed (though not unrelaxed, see below), and radiation damage is much slower, if the photon energy remains below about 30 eV [9]. The observation of  $k$ -dispersive spectral features in ARPES, signaling the high and stable long-range order of TTF-TCNQ surfaces, was first reported by Zwick *et al.* [10]. In fact, to our knowledge TTF-TCNQ has up to now remained the only charge-transfer salt where this has been possible.

### 3 Experimental Details

The ARPES experiments have been performed both at our home laboratory with He I radiation from a discharge lamp and at BESSY (Berlin) using synchrotron radiation. Energy and angular resolution amounted to 60 meV and  $\pm 1^\circ$ , respectively. All data were taken at a base pressure below  $10^{-10}$  mbar and at temperatures of 60 K or higher, *i.e.* above the Peierls transition. Single crystals with typical dimensions of  $2 \times 5 \times 0.2$  mm<sup>3</sup> (long axis along the b-direction) were cleaved *in situ* parallel to the a-b plane to obtain clean surfaces. The chemical and structural quality of these surfaces was characterized by x-ray photoemission (XPS) and scanning tunneling microscopy (STM). All ARPES data shown here have been taken within a time window of 2-3 hours after cleavage and before noticeable radiation damage occurred [9].

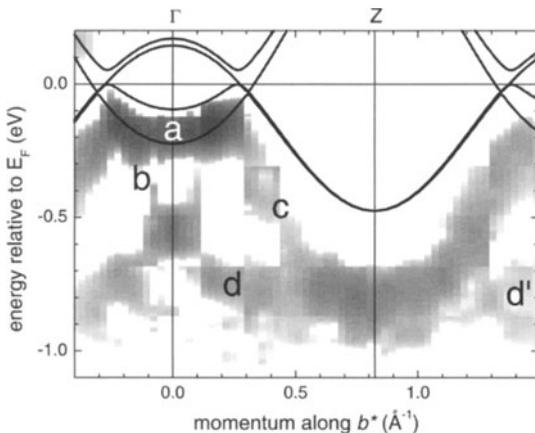


**Fig. 2.** ARPES dispersions of TTF-TCNQ measured parallel (left) and perpendicular (right) to the molecular stacking direction

#### 4 Results and Discussion

Figure 2 shows ARPES spectra along the  $\Gamma$ -Z and  $\Gamma$ -Y directions of the Brillouin zone (cf. Fig. 1), *i.e.* with wavevector  $k$  varying along and perpendicular to the 1D stacking axis, respectively. While parallel to the 1D axis the spectral conduction band features display a pronounced energy-*vs.*-wavevector dependence, virtually no variations are observed in the perpendicular direction. This directly reflects the strongly anisotropic  $k$ -space dispersion of the conduction band states, consistent with the huge anisotropy in the conductivity.

A gray-scale representation of the ARPES data along the 1D axis, which allows to trace the dispersions more easily, is given in Fig. 3. Also shown are the conduction band dispersions obtained from a density-functional band calculation using the generalized gradient approximation (GGA) [5]. According to the GGA calculation the bands occupied at the  $\Gamma$ -point are derived from the partially populated TCNQ LUMO, whereas the bands occupied at Z have their origin in the correspondingly “hole-doped” TTF HOMO. Both types of bands come as doublets, because the unit cell contains two TTF and two TCNQ stacks each. Due to a small interaction between neighboring TCNQ stacks their band degeneracy is slightly lifted, whereas it remains largely preserved for the TTF bands. The position of the common Fermi vector of both band complexes is determined by the charge transfer between TTF and TCNQ stacks.

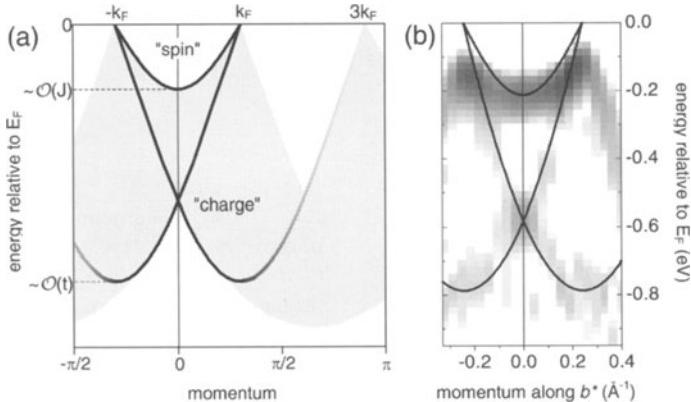


**Fig. 3.** Gray-scale plot of the ARPES dispersions (displayed as negative second energy derivative of the intensity,  $-\frac{d^2 I}{dE^2}$ ) in comparison to the conduction band dispersions obtained by a density-functional band calculation [5]

The comparison of experimental and GGA dispersions reveal a number of qualitative and quantitative discrepancies. To begin with, the theoretical TTF band has an experimental counterpart (labeled **c** in Figs. 2 and 3), which displays the same U-shaped dispersion but is characterized by a width almost twice as large. In the  $k$ -range around the zone center, where band theory predicts the TCNQ band to be occupied, we find three experimental features (**a**, **b**, and **d**). Again, two of them (**a** and **b**) could be related to the GGA bands by rescaling their band width by a factor of  $\sim 2$ . Keeping in mind the extreme surface sensitivity of ARPES the enhanced band widths may result from a structural relaxation of the TCNQ and TTF stacks lying in the topmost molecular layers. More specifically, it seems plausible that the tilt angle of the molecules with respect to the 1D axis (cf. Fig. 1), which results from the competition between maximum covalent bonding along the stacks and minimum electrostatic energy (all molecules are ionic due to the charge transfer), is modified for the surface molecules as they experience a different Madelung potential. This would imply a notable renormalization of the nearest neighbor hopping integrals and hence of the corresponding band widths.

However, the band calculation offers no explanation for the origin of experimental feature **d** (and its replica **d'** in the next Brillouin zone). If it were an additional surface state (for which there is no obvious explanation), its occupancy throughout the Brillouin zone would severely offset the delicate charge balance between the TTF and TCNQ conduction bands and should thus lead to a pronounced shift of the Fermi vector along the 1D axis compared to its position in the bulk. This is obviously not the case.

The failure of band theory is consistent with other experimental indications that the electronic properties of TTF-TCNQ are dominated by strong correlation effects [6]. Therefore a 1D Hubbard model description seems more appropriate. Here the delocalization of the charge carriers is described by an



**Fig. 4.** (a) Schematic electron removal spectrum of the doped 1D Hubbard model with band filling  $1/2 < n < 2/3$  (see text for details). (b) Theoretical spin and charge branch dispersions of the 1D Hubbard model calculated for  $U = 1.96$  eV,  $t = 0.4$  eV, and  $n = 0.59$  compared to the ARPES dispersion of the TCNQ-derived conduction band complex

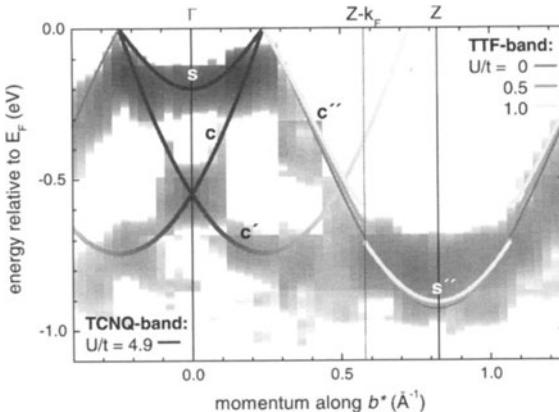
intermolecular hopping integral  $t$  whereas the localizing effect of the Coulomb repulsion is captured by an *intramolecular* interaction energy  $U$ . While early treatments of the model had been restricted to the strong coupling limit ( $U/t \rightarrow \infty$ ), there has recently been great progress to calculate detailed spectral properties also for finite coupling strengths ( $U/t \sim 1$ ) [2,3]. A schematic view of the  $k$ -dependent single-particle spectrum below the chemical potential (*i.e.*, the electron removal spectrum) is given in Fig. 4(a). In a simple picture a hole generated by removal of an electron (by e.g. photo-excitation) decomposes (or “fractionalizes”) into decoupled spin and charge excitations (“spinons” and “holons”). As energy and momentum of the original hole can be distributed among these collective modes, a broad excitation continuum results (shaded area in the figure). However, due to the phase space available for hole fractionalization the continuum is not structureless. Rather, it is dominated by  $k$ -dispersive singularities (solid curves in Fig. 4(a)) which correspond to situations, in which either the holon propagates with the spinon having zero energy, or *vice versa*. The resulting “charge” and “spin” branches scale with characteristic energies  $t$  and  $J$ , respectively, where the latter is the effective magnetic exchange energy [ $J = \frac{2t^2}{U}(n - \frac{\sin(2\pi n)}{2\pi})$ ,  $n$  denotes the band filling]. The loss of spectral weight in the charge branch for  $k$  between  $k_F$  and  $3k_F$ , which was already suggested by strong-coupling calculations, has recently been confirmed by the dynamical density-matrix renormalization group (DDMRG) method to be a generic property of the model [3].

Comparing the topology of the charge and spin branches to the TCNQ-related ARPES dispersions in Fig. 3 one observes striking similarities. A direct comparison between experimental TCNQ dispersions and a 1D Hubbard

model calculation using the Bethe ansatz is contained in Fig. 4(b) [5]. In the calculation the band filling was fixed at  $n = 0.59$  (consistent with the charge transfer) and  $U$  and  $t$  were chosen such to obtain optimum agreement with experiment. Indeed, the Hubbard model provides an excellent description of the data for  $t = 0.4$  eV and  $U = 1.96$  eV, corresponding to a rather moderate coupling strength of  $U/t = 4.9$  in the TCNQ stacks. The experimental feature **a** is thus identified as the “spin” branch of the Hubbard model, whereas structures **b** and **d** constitute the “charge” branch which together with its  $k \rightarrow -k$  counterpart forms the characteristic X-shaped dispersion in Fig. 4. While the resulting Coulomb energy  $U$  is consistent with earlier estimates, the hopping rate  $t$  would imply a bare band width  $W = 4t$  which is doubled with respect to band theory and its value derived from bulk properties [6]. We attribute this again to a surface relaxation of the topmost molecule stacks. For the volume electronic structure we can estimate a corresponding  $U/t$  ratio by assuming that  $U$  as an intramolecular property does not vary much between surface and bulk and by deriving a bulk value for  $t$  from the GGA band width. This yields  $U/t \approx 11$  for the bulk. From these parameters we can also infer a magnetic energy scale  $J \approx 21$  meV, which agrees favorably with experimental values ranging between 17 and 30 meV derived from magnetic susceptibility measurements [11,12].

In contrast to the TCNQ band the ARPES dispersion of the TTF band shows no unusual behavior except for the enhanced band width. According to the GGA calculation this band is more than half-filled ( $n = 2 - 0.59 > 1$ ). For such a situation the 1D Hubbard model would predict a gap at wavevector  $k = \pi - k_F$  in the dispersion, separating a spin- and a charge-like branch from each other [13]. Its size depends exponentially on  $U/t$ . However, the ARPES data give no evidence for such a gap within experimental resolution (cf. Fig. 5), leading to an upper limit of  $U/t < 0.2(0.4)$  for the surface (volume) and thus implying only weak correlation effects. While this finding is consistent with a temperature-independent (Pauli-like) spin susceptibility of the TTF stacks [14], it seems to be at variance with the observation of  $4k_F$  charge density fluctuations which had been assigned to these stacks [6,15]. On the other hand,  $4k_F$  fluctuations are a signature of *long-range* Coulomb interactions not included in our bare Hubbard model fit which accounts only for on-site interactions. The additional consideration of non-local Coulomb effects by adding at least a nearest neighbor interaction term severely complicates the problem. The calculation of reliable spectral functions for such an extended Hubbard model is unfortunately still out of reach. Preliminary results from exact diagonalization of small finite systems [16] suggest however that the effect of long-range interaction on the spectral dispersion is not as dramatic as would be required to resolve the TTF puzzle. This calls strongly for further theoretical work on the extended Hubbard model.

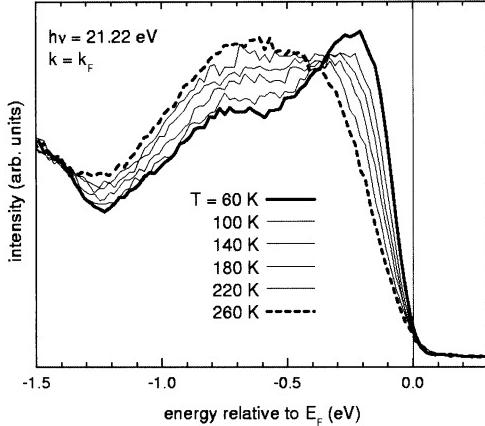
In return, the absence of  $4k_F$  fluctuations on the TCNQ stacks [6] justifies the use of a purely local Hubbard model for them. The strong 1D correlation



**Fig. 5.** Hubbard model fits to the TCNQ- and TTF-derived ARPES dispersions. The spin and charge branches of the model are assigned by the labels  $s, s''$  and  $c, c', c''$ , respectively

effects observed in the experimental TCNQ dispersions are further corroborated by a very unusual temperature dependence of the ARPES spectra [4,5] which is shown in Fig. 6. With increasing temperature from 60 K to nearly room temperature the data reveal a dramatic transfer of intensity from low to high excitation energies on an energy scale of the total band width. The effect is fully reversible and conserves the integrated spectral weight within experimental accuracy. A weight transfer over an energy so much larger than the thermal energy scale cannot be explained by conventional electron-phonon coupling or possible Peierls fluctuations. Rather, it is a signature of electronic correlation effects and consistent with the temperature dependence calculated for the 1D  $tJ$  model in the limit  $J/t \rightarrow 0$  (corresponding to the  $U/t \rightarrow \infty$  case of the Hubbard model) [17]. There it was found that, compared to the zero temperature spectrum, considerable spectral weight is redistributed from the spinon peak at the Fermi level to the bottom of the holon band already for low temperatures  $0 < k_B T \ll t$ , exactly as observed in our data. We expect this result to hold qualitatively also for finite  $J/t$  or  $U/t$ .

We finally turn to the behavior of the spectra close to the chemical potential. Theoretically, the low-energy properties of the 1D Hubbard model follow the TLL phenomenology, which is parameterized by the charge stiffness constant  $K_\rho$  [1]. For example, the low-energy onset of the  $k$ -integrated electron removal spectrum is expected to follow a  $|\omega|^\alpha$  power law with  $\alpha = \frac{1}{4}(K_\rho + K_\rho^{-1} - 2)$ . From the Hubbard model parameters relevant for the TCNQ band one obtains  $K_\rho \approx 0.68$  [3,18] yielding  $\alpha \approx 0.038$ . With such an exponent the spectral onset would practically be indistinguishable from a conventional metallic Fermi edge. The  $k$ -resolved spectrum at the Fermi vector  $k_F$  would even display singular behavior ( $\propto |\omega|^{\alpha-1}$ ) [1]. Quite contrary to this theoretical expectation the experimental  $k_F$  spectrum displays an almost linear



**Fig. 6.** Temperature dependence of the ARPES spectrum at  $k = k_F$

energy dependence within the first 200 meV of the chemical potential (see, *e.g.*, Fig. 6), which is also seen in the  $k$ -integrated spectrum [5]. This indicates that the simple 1D Hubbard model which works so well for the finite energy spectral properties fails at low excitation energies. There are various possible explanations for this finding. First of all, the physics of the 1D Hubbard model is strictly applicable only for energies larger than the transverse hopping integral *between* the 1D stacks. As this is small but non-zero it would be not surprising to observe deviations from canonical 1D TLL behavior at low energies. Another explanation is based on the effect of long-range Coulomb interaction beyond a purely local Hubbard model, which may lead to larger power law exponents of up to  $\alpha \sim 1$  [19]. It has been argued that such values for  $\alpha$  may also result from the inevitable existence of impurities and/or defects on the 1D stacks, which localize the conduction electrons to segments of finite length and thus modify their description within the concept of “bounded Luttinger liquids” [20].

Furthermore, the failure of the Hubbard model at low energies may not have a purely electronic origin. In fact, from the very existence of the Peierls instability we know that electron-phonon coupling also plays an important role in TTF-TCNQ. However, a simple interpretation of the observed linear spectral onset in terms of a Peierls pseudo gap induced by CDW fluctuations can be ruled out, as the characteristic energy scale given by the low-temperature Peierls gap is only 20 meV [6] and hence much smaller than the range of the linear onset. This could indicate strong coupling to other phonons than those involved in the Peierls transition and is consistent with the phonon spectrum of TTF-TCNQ which indeed reaches up to  $\sim 200$  meV [21]. A detailed understanding of the low-energy spectral properties of TTF-TCNQ will thus require the consideration of electronic correlations *and* electron-phonon coupling effects on an equal footing. Unfortunately, no such theory exists as yet.

## 5 Conclusions

In summary, the electronic structure of the quasi-1D organic conductor TTF-TCNQ as probed by ARPES deviates strongly from a conventional band structure picture. When accounted for the effects of a possible surface relaxation, the finite energy spectral dispersions can be consistently and even quantitatively explained by the single-particle spectrum of the 1D Hubbard model. This interpretation is further supported by a temperature-dependent redistribution of spectral weight over energies much larger than the thermal energy. In particular the TCNQ-derived conduction band thus displays spectroscopic signatures of spin-charge separation on an energy scale of the band width. Open questions concern the low  $U/t$  ratio found for the TTF band and the low-energy spectral behavior, which points to the possible additional importance of electron-phonon interaction, long-range Coulomb effects, or three-dimensional coupling.

These results highlight the great importance of electronic correlations and other many-body effects in molecular organic compounds. In fact, the delicate balance between intramolecular Coulomb interaction and intermolecular electron hopping is a typical trademark of these materials and at the heart of their complex electronic phase diagram. Here we have studied the additional role of reduced dimensionality, and it may be tempting to speculate that the physics encountered here may also be relevant for quasi-1D biological molecules such as DNA.

## Acknowledgements

We wish to thank U. Schwingenschlögl for experimental support and M. Dressel (Stuttgart) and C.S. Jacobsen (Lyngby) for providing the TTF-TCNQ crystals. For their theoretical contributions we are indebted to P. Blaha (Vienna) and J.M.P. Carmelo (Braga). This work was supported by the Deutsche Forschungsgemeinschaft (CL 124/3 and SFB 484).

## References

1. J. Voit, Rep. Prog. Phys. **58**, 977 (1995).
2. J.M.P. Carmelo *et al.*, cond-mat/0305568; Nucl. Phys. B **683**, 387 (2004); cond-mat/0303279; Phys. Rev. B **68**, 085104 (2003).
3. H. Benthien, F. Gebhard, and E. Jeckelmann, cond-mat/0402664.
4. R. Claessen *et al.*, Phys. Rev. Lett. **88**, 096402 (2002).
5. M. Sing *et al.*, Phys. Rev. B **68**, 125111 (2003).
6. S. Kagoshima, H. Nagasawa, and T. Sambongi: *One-Dimensional Conductors* (Springer, Berlin 1988).
7. A. Schwartz *et al.*, Phys. Rev. B **58**, 1261 (1998).
8. T. Lorenz *et al.*, Nature **418**, 614 (2002).
9. M. Sing *et al.*, Phys. Rev. B **67**, 125402 (2003).

10. F. Zwick *et al.*, Phys. Rev. Lett. **81**, 2974 (1998).
11. J.B. Torrance, Y. Tomkiewicz, and B.D. Silverman, Phys. Rev. B **15**, 4738 (1977).
12. S. Klotz *et al.*, Phys. Rev. B **38**, 5878 (1988).
13. R. Claessen *et al.*, J. Phys. IV, **144**, 51 (2004).
14. T. Takahashi *et al.*, J. Phys. C: Solid State Phys. **17**, 3777 (1984).
15. J.P. Pouget, in *Semiconductors and Semimetals*, Vol. **37**, E. Conwell (Ed.) (Academic Press, New York 1988) p. 87.
16. S. Yunoki, T. Tohyama, and S. Maekawa, Physica B **230-232**, 1050 (1997).
17. K. Penc and M. Serhan, Phys. Rev. B **56**, 6555 (1997).
18. H.J. Schulz, Phys. Rev. Lett. **64**, 2831 (1990).
19. A.K. Zhuravlev and M.I. Katsnelson, Phys. Rev. B **64**, 033102 (2001).
20. J. Voit, Y. Wang, and M. Grioni, Phys. Rev. B **61**, 7930 (2000).
21. J.E. Eldridge, Solid State Commun. **105**, 427 (1998).

# Quantum Critical Metals: Beyond the Order Parameter Fluctuations

Qimiao Si

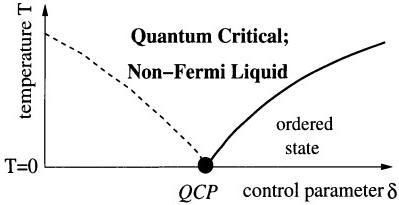
Department of Physics & Astronomy, Rice University  
Houston, TX 77005, USA

**Abstract.** The standard description of quantum critical points takes into account only fluctuations of the order parameter, and treats quantum fluctuations as extra dimensions of classical fluctuations. This picture can break down in a qualitative fashion in quantum critical metals: non-Fermi liquid electronic excitations are formed precisely at the quantum critical point and appear as a part of the quantum-critical spectrum. In the case of heavy fermion metals, it has been proposed that the non-Fermi liquid behavior is characterized by the destruction of a Kondo effect. The latter invalidates Hertz's Gaussian theory of paramagnons and leads to an interacting theory that is "locally quantum critical". We summarize the theoretical and experimental developments on the subject. We also discuss their broader implications, and make contact with recent work on quantum critical magnets.

## 1 The Order Parameter Fluctuation Theory of Quantum Critical Points and Its Breakdown

Phase transitions come in different varieties. Generically, they are characterized by the onset of an order parameter. A classical critical point, occurring at a finite temperature phase transition of second-order, is described in terms of a coarse-grained theory of spatial, but time-independent, fluctuations of the order parameter [1]. Such a description also serves as the basis to categorize the universality classes of critical points. Quantum critical points (QCPs) take place at zero temperature. They differ from their classical counterparts in that the static (classical) and dynamic (quantum) fluctuations are mixed and both have to be incorporated in the critical theory. It is, however, standard to assume that they too can be described in terms of fluctuations of the order parameter: the only distinction being that the fluctuations are not only in space but also in (imaginary) time [2,3].

It has been realized over the past few years that this picture can break down in a qualitative fashion in quantum critical metals [4,5]. Non-Fermi liquid excitations emerge precisely at the QCP, and they need to be kept as a part of the quantum-critical spectrum. This is illustrated in Fig. 1. On the one hand, quantum criticality is the mechanism for the non-Fermi liquid behavior. On the other hand, the non-Fermi liquid excitations feed back and change the universality class of the underlying QCP. The experimental motivations have largely come from heavy fermion QCPs [6,4,5]. Discussions of a similar spirit can be found in an earlier pedagogical article [7].



**Fig. 1.** Schematics of quantum phase transitions in strongly correlated metals. Quantum criticality leads to non-Fermi liquid excitations which, in turn, become a part of the quantum-critical spectrum and give rise to new types of quantum-critical point

The notion that quantum fluctuations at a second-order zero-temperature phase transition amount to adding extra dimensions of purely classical fluctuations dates back to the work on the Ising model in a transverse field [8,3]. Within the renormalization group framework, Hertz [2] carried out such a mapping for a model of a metal undergoing a quantum spin-density-wave (SDW) transition. He showed that the number of the extra dimensions is equal to the dynamic exponent  $z$ . The result [2,9] is a  $d_{eff} = d + z$  (where  $d$  is the spatial dimensionality) dimensional critical theory of paramagnons [10,11,12,13]. It was not until recent years that quantum critical metals became experimentally realized, mostly in heavy fermion metals. In many cases, the experiments contradict the Hertz picture [6,4,5] (see below).

From a theoretical point of view, the possible breakdown of this picture can be seen in a number of ways. First, the construction of the order parameter fluctuation theory proceeds by integrating out fermions altogether. However, non-Fermi liquid excitations emerge exactly at the QCP (Fig. 1); they go away inside the phases on both sides of the QCP. These emergent excitations should be considered as a part of the quantum-critical spectrum. The resulting critical theory is then more than just fluctuations of the order parameter.

Second, in the mapping of quantum fluctuations into extra dimensions of classical fluctuations, one writes the partition function of a quantum mechanical system in terms of a summation over configurations of classical variables in space and time:

$$Z \sim \sum_{\text{config in } (\mathbf{x}, \tau)} Z(\text{config}). \quad (1)$$

However, if the partition function for the individual configurations in space and time is not positive semidefinite, Eq. (1) would not necessarily correspond to a classical statistical mechanical theory in  $d + z$  dimensions. For electronic systems, this has been known through the “fermion sign problem” encountered in quantum Monte-Carlo methods [14]. The sign problem also appears in quantum spin systems. Indeed, recent work has demonstrated its effect in QCPs of two-dimensional quantum magnets [15] (see below).

## 2 Quantum Critical Heavy Fermions

Quantum criticality is of particular interest to the physics of strongly correlated metals. It provides a route towards non-Fermi liquid behavior [16,17], and may also serve as a means of generating collective states such as unconventional superconductors [18,19]. Heavy fermions have been playing an especially important role, for the simple reason that here QCPs have been explicitly identified [6].

The contradiction to the paramagnon description was initially found in the inelastic neutron scattering experiments of heavy fermion metals near an antiferromagnetic QCP [20,21,22]. The prediction of the paramagnon theory is straightforward. Due to Landau damping, the dynamic exponent  $z = 2$  in the antiferromagnetic case. The effective dimensionality of the fluctuations of the paramagnons,  $d_{\text{eff}} = d + 2$ , is either larger than or equal [21] to 4, the upper critical dimension. The fixed point is Gaussian, and the frequency dependence of the dynamical spin susceptibility is simply given by the Landau damping [2], with an exponent 1. At the same time, its temperature dependence is controlled by the quartic coupling among paramagnons; this coupling is dangerously irrelevant and produces a temperature exponent that is larger than 1 [9]. An important corollary is that the dynamical spin susceptibility necessarily violates  $\omega/T$  scaling [3]. (In special cases, integrating out fermions within the Hertz framework makes the coupling constants of the paramagnon theory non-analytic [23,24]. Such effects may lead to fractional exponents in the dynamical spin susceptibility at  $\mathbf{q} \sim \mathbf{Q}$ , but not  $\omega/T$  scaling [24].)

The experiments [20,21,22], on the other hand, observe  $\omega/T$  scaling. Moreover, the critical exponent for the frequency and temperature dependences is fractional. Finally, this same fractional exponent is observed even at wavevectors far away from the antiferromagnetic ordering wavevector  $\mathbf{Q}$ , in a form [20]

$$\chi(\mathbf{q}, \omega) = \frac{\text{const.}}{f(\mathbf{q}) + (-i\omega)^\alpha W(\omega/T)}, \quad (2)$$

where  $f(\mathbf{q})$  vanishes as  $\mathbf{q}$  approaches the antiferromagnetic wavevector  $\mathbf{Q}$  and stays non-zero elsewhere. The fact that this same fractional exponent  $\alpha$  appears essentially everywhere in the Brillouin zone suggests that its origin is local, as noted in Ref. [20].

Heavy fermions near a magnetic quantum phase transition should be well-described in terms of a Kondo lattice model:

$$\mathcal{H} = \sum_{ij,\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + \sum_i J_K \mathbf{S}_i \cdot \mathbf{s}_{c,i} + \sum_{ij} I_{ij} \mathbf{S}_i \cdot \mathbf{S}_j. \quad (3)$$

We are considering metallic systems, away from half-filling (*i.e.*, the number of conduction electrons is other than one per  $f$  electron). We can define the tuning parameter of this model as  $\delta$ , the ratio of the RKKY interaction to the bare Kondo scale  $T_K^0$ . For negligible  $\delta$ , the Kondo effect dominates [25,26]: the local moments are quenched and the excitations below some energy scale,

$E_{\text{loc}}^*$ , are spin- $\frac{1}{2}$  and charge- $e$  Kondo resonances [27]. The result is a paramagnetic heavy fermion metal. For large  $\delta$ , on the other hand, the interactions between the local moments play the dominant role. For dimensions higher than one and in the absence of strong geometrical frustration, the ground state is expected to be magnetically ordered. Typically, this would be an antiferromagnetic metal (as can be expected when the system is not too far away from half-filling).

What is the nature of the quantum phase transitions? It was suggested in Ref. [28] that the behavior of  $E_{\text{loc}}^*$  can be used as a characterization. As defined earlier, this is the scale below which local-moments are turned into Kondo resonances. It can also be operationally defined in terms of the spin damping or “spin self-energy”  $M(\omega)$  (see below):  $E_{\text{loc}}^*$  separates the region ( $\omega \gg E_{\text{loc}}^*$ ) where  $M(\omega)$  contains a fractional exponent and that ( $\omega \ll E_{\text{loc}}^*$ ) where it is linear in frequency. Two different classes of QCP occur depending on whether  $E_{\text{loc}}^*$  stays finite or vanishes at the QCP [28]. Subsequent microscopic works [4,29,30,31] have indeed shown such QCPs. Related approaches have also been developed in Refs. [32,5,33].

### 3 Microscopic Results and Their Robustness

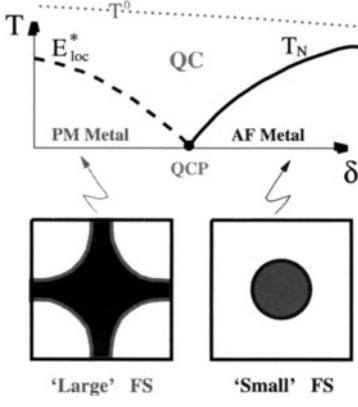
#### 3.1 Microscopic Results

The microscopic analysis was carried out within the extended dynamical mean field theory of Kondo lattice systems [29,34] (this approach goes beyond the dynamical mean field theory [35,36]; for the initial development of the method, see Refs. [37,38]). In this approach, the spin spin-energy is taken to be momentum-independent, which is expected to be valid for antiferromagnetic quantum critical metals provided the spatial anomalous dimension  $\eta = 0$ .

The “local quantum critical point” arises in this microscopic approach to the Kondo lattice model. The Kondo coupling is relevant in the renormalization group sense on the paramagnetic side, leading to Kondo singlet formation and Kondo resonances below a finite energy scale  $E_{\text{loc}}^*$ . At the QCP, however, the Kondo coupling is on the verge of becoming irrelevant. So  $E_{\text{loc}}^*$  goes to zero precisely at the QCP; see Fig. 2.

The microscopic result also provides yet another alternative means to define  $E_{\text{loc}}^*$ , which appears as an infrared cutoff scale in the local spin susceptibility,  $\chi_{\text{loc}} \equiv <\chi(\mathbf{q}, \omega)>_{\mathbf{q}}$ . On the paramagnetic metal side,  $E_{\text{loc}}^*$  is finite, and  $\chi_{\text{loc}}$  has the Pauli behavior reflecting the Kondo resonances. At the QCP, on the other hand,  $E_{\text{loc}}^* = 0$  and  $\chi_{\text{loc}}$  is singular. The singularity is however weaker than the Curie behavior in the non-scaling regime above the ultraviolet cutoff scale  $T^0$ . ( $T^0$  is of the order of the single-ion Kondo temperature  $T_K^0$ , and is always finite.)

The initial analytic result for the quantum-critical dynamics was derived with the aid of an  $\epsilon$ -expansion of the self-consistent Bose-Fermi Kondo



**Fig. 2.** Local quantum critical point. The Fermi surface undergoes a sudden reconstruction across the QCP; within the paramagnetic Brillouin zone, it is ‘large’ on the heavy fermion side and ‘small’ on the antiferromagnetic side

model [4,29]. The local spin susceptibility is

$$\chi_{\text{loc}}(\omega, T = 0) = \frac{1}{2\Lambda} \ln \frac{\Lambda}{-i\omega}, \quad (4)$$

where the energy scale  $\Lambda \approx T_K^0$ . The corresponding spin self-energy has the form

$$M(\omega, T) \approx -I_Q + A (-i\omega)^\alpha W\left(\frac{\omega}{T}\right), \quad (5)$$

and the exponent is

$$\alpha = \frac{1}{2\rho_i(I_Q)\Lambda}. \quad (6)$$

Since the product  $\rho_i(I_Q)\Lambda$  is expected to be of order unity at the transition [4,29],  $\alpha$  should be fractional.

The singular fluctuations seen in the local susceptibility reflects the embedding of the criticality of the local Kondo physics in the criticality associated with the antiferromagnetic ordering. The singular form might raise the concern that the local quantum critical point is pre-empted by magnetic ordering, which would turn the transition first order. It was argued that a second order transition is natural: for two-dimensional magnetic fluctuations, a divergent local susceptibility necessarily accompanies a divergent peak susceptibility (provided the spatial anomalous dimension  $\eta = 0$ , which is to be expected at the quantum transition [29]).

This issue has been further addressed numerically [30]. We’ve considered the model with Ising anisotropy, the case with a maximum tendency towards magnetic ordering. Indeed, the extrapolated zero-temperature transition is second order. The extended dynamical mean field analysis turns out to be somewhat subtle. The Bose-Fermi Kondo model with Ising anisotropy and in the absence of a static local field has a quantum phase transition: while it is in a paramagnetic Kondo state when the coupling  $g$  to the bosonic bath is weak,

the phase at large  $g$  contains a finite Curie constant [39]. Since the existence of a Curie constant is hard to see at high temperatures, it is important to reach sufficiently low temperatures in order to determine the nature of the zero-temperature phase transition, as was done in Ref. [30]. The continuous nature was not seen in a numerical study at higher temperatures [40].

The quantum Monte-Carlo access to the local quantum critical point also provides an opportunity to numerically determine the exponent  $\alpha$  defined in Eqs. (5,6). In the Ising case, it is found [31] to be approximately 0.72. Two observations about the fractional exponent are worth making. First, the exponent turns out to be nearly universal for this model: within the numerical accuracy, the value is the same for four different sets of initial parameters. Second, there is limitation to an analytic study based on a rotor O( $N$ ) generalization of the model [the physical case, studied in the numerical work, corresponds to  $N = 1$ ]. While it captures the local quantum criticality (destruction of the Kondo effect at the magnetic QCP), the rotor large- $N$  generalization, in either the leading order [31] in  $1/N$  or the next-to-leading order [41], fails to capture the fractional exponent shown in the  $N = 1$  case. This limitation is understood to be the result of a “pinning” of the critical amplitude of the local susceptibility in the large- $N$  limit [31].

### 3.2 Beyond Microscopics

These results have been argued [4,29] to be robust beyond the microscopic approach, when the exponent  $\alpha$  is fractional. The key point is that, for the long-wavelength spin fluctuations, the dynamic exponent  $z = 2/\alpha$ . For  $\alpha < 1$ , it is internally consistent to have the spatial anomalous dimension  $\eta = 0$ . The temporal fluctuations, on the other hand, contain a large anomalous dimension reflecting the destruction of Kondo effect.

### 3.3 In What Sense Is the QCP Local?

There are several ways to characterize the local nature of the QCP.

First, there is a localization of  $f$ -electrons at the QCP. On the paramagnetic side, the  $f$ -local moments are part of the low-energy electronic excitations<sup>1</sup>. This happens as a result of the Kondo effect: through the Kondo singlet formation, the local moments and conduction electrons are entangled and the initially charge-neutral  $f$ -moments are turned into charge- $e$  and spin- $\frac{1}{2}$  quasiparticle excitations. In other words, the  $f$ -moments behave essentially as delocalized electrons. Indeed, the Fermi surface is adiabatically connected to that of a system in which the  $f$ -electrons are simply taken as non-interacting electrons! In the case of one  $f$ -element per unit cell, the

---

<sup>1</sup> Here we are concerned with the low-energy behavior. At high-energy atomic scales, there are no local moments and  $f$ -electrons are always a part of the electronic spectrum.

Fermi volume is proportional to  $1 + x$ , where 1 counts the  $f$ -local moment and  $x$  is the number of conduction electrons per unit cell. On the magnetically ordered side, on the other hand, the  $f$ -moments stay charge neutral and are not a part of the electron fluid. The Fermi surface is that of the conduction electrons alone, under the influence of a static and periodic magnetic field produced by the antiferromagnetic order parameter. Even in the paramagnetic Brillouin zone, the Fermi volume is proportional to  $x$ . The discontinuous reconstruction of the Fermi surface takes place precisely at the QCP; see Fig. 2. Exactly at the QCP, the effective electronic mass diverges over the *entire* Fermi surface.

Second, the local nature is also reflected in the magnetic dynamics. The same anomalous exponent in the frequency and temperature dependences appears essentially *everywhere* in the Brillouin zone. And the local spin susceptibility is singular.

Both characterizations are symptomatic of the embedding of the destruction of Kondo effect into the magnetic ordering transition. The non-Fermi liquid excitations, which arise when the weight of the Kondo resonance just goes to zero, are part of the quantum-critical spectrum. The critical theory must incorporate these non-Fermi liquid excitations, and so it is no longer just the  $(d+z)$ -dimensional fluctuations of the order parameter as in the paramagnon theory of the  $T = 0$  SDW transition.

## 4 Experiments

### 4.1 Spin Dynamics

The local quantum critical point is an interacting fixed point, in contrast to the Gaussian fixed point of the paramagnon theory. This is directly manifested in the spin dynamics. The  $\omega/T$  scaling and fractional exponent of the dynamical spin susceptibility have, as mentioned earlier, already been found in  $\text{CeCu}_{6-x}\text{Au}_x$ . In addition, the spin self-energy shown in Eq. (5) yields a dynamical susceptibility of the form given in Eq. (2), with the same fractional exponent appearing in a broad region of the Brillouin zone. (Closely-related features in the dynamics [42,43] and thermodynamics [44] appear in the compound  $\text{UCu}_{5-x}\text{Pd}_x$ .)

The static bulk spin susceptibility is also expected to show the same fractional exponent<sup>2</sup>:

$$\chi(T) = \frac{1}{\Theta + BT^\alpha} \quad (7)$$

This was already seen in  $\text{CeCu}_{6-x}\text{Au}_x$  early on [22]. Within a more limited temperature range, a fractional exponent has also been observed in

---

<sup>2</sup> If the total spin is conserved, the spin conservation law may modify the behavior of the susceptibility near  $\mathbf{q} = \mathbf{0}$ .

$\text{YbRh}_2\text{Si}_2$  [45]. Eq. (7) is also compatible with the data [46,47] in the normal state of  $\text{CeCoIn}_5$ , although the extent to which this system is quantum critical remains to be established.

The NMR relaxation rate probes local spin dynamics, at a frequency that is much smaller than the typical measurement temperature. Assuming a featureless hyperfine-coupling, the relaxation rate was predicted to contain a temperature-independent component [4,29]:

$$\frac{1}{T_1} \sim A_{hf}^2 \frac{\pi}{8A} \quad (8)$$

This has been subsequently observed in the Si-site NMR relaxation rate of  $\text{YbRh}_2\text{Si}_2$  [48]. For  $\text{CeCu}_{6-x}\text{Au}_x$ , recent measurement [49] of the Cu-site NMR sees  $1/T_1 \sim T^\alpha$  instead. It seems that the only way this result can be compatible with the inelastic neutron-scattering data is to invoke a strongly momentum-dependent hyperfine coupling constant such that the NMR relaxation rate is dominated by contributions from generic wavevectors. If that is indeed the case, the NMR experiment would provide additional evidence that the very same fractional exponent appears at generic wavevectors.

## 4.2 Thermodynamics

A divergent specific heat coefficient (specific heat divided by temperature) at the QCP is expected from the localization of the  $f$ -electrons: the effective mass diverges over the entire Fermi surface, as mentioned earlier. However, a divergent specific heat coefficient can also arise in some special cases [50] of the SDW quantum critical point, due to contributions of the so-called “hot spots” alone.

A more definite probe turns out to come from thermodynamic ratios. Recall that, at classical critical points, thermodynamic quantities such as specific heat diverge and provide the means to measure scaling exponents. Since a QCP occurs at zero-temperature, the third-law of thermodynamics dictates that the specific heat has to go to zero. On the other hand, thermodynamic ratios can still diverge. A practical example is the Grüneisen ratio – the ratio of the thermal expansion,  $\alpha \equiv \frac{1}{V} \frac{\partial V}{\partial T}$ , to the specific heat,  $c_p$ :

$$\Gamma = \frac{\alpha}{c_p} \sim \frac{\partial S / \partial p}{T \partial S / \partial T}. \quad (9)$$

Under scaling, this ratio must diverge at a QCP [51]. Moreover, the Grüneisen temperature exponent  $x$ , as defined by  $\Gamma_{\text{crit}}(T) \sim 1/T^x$ , measures the scaling dimension of the most singular operator coupled to pressure. It turns out that, within the paramagnon theory of an antiferromagnetic SDW transition,  $x$  is equal to 1 (up to logarithmic corrections in some cases). At a local quantum critical point, on the other hand,  $x$  can be fractional.

This divergence has now been observed experimentally [52] in two heavy fermion compounds,  $\text{YbRh}_2\text{Si}_2$  and  $\text{CeNi}_2\text{Ge}_2$ . For  $\text{CeNi}_2\text{Ge}_2$ , the Grüneisen

temperature exponent is found to be equal to 1. For  $\text{YbRh}_2\text{Si}_2$ , it is fractional and is approximately 0.7 – this value is inconsistent with the paramagnon theory but is compatible with the scaling dimension (0.66 to the second order of the  $\epsilon$ -expansion) of the operator that tunes the system to local quantum criticality.

### 4.3 Electronic Measurements

The dynamical and thermodynamic measurements provide compelling evidence for the breakdown of the paramagnon theory in some of the heavy fermion metals, and make a strong case for the local quantum critical picture. Ultimately, one would like to test the destruction of the Kondo effect directly, and this can only be done through electronic measurements. Such evidence is just emerging, from for example the recent Hall measurements [53].

## 5 Broader Context

Heavy fermions represent a prototype strongly correlated system to study quantum critical physics. It is simpler than the doped Mott insulators in that a large separation of energy scales exists: the local-moments are well defined over a broad energy range, and the Kondo coupling is much smaller than both the bare conduction electron bandwidth and the atomic Coulomb interactions. It nonetheless is similar to the Mott-Hubbard systems in that the strong Coulomb interactions lead to a microscopic Coulomb blockade. The resulting projection of Hilbert space is essential to the destruction of the Kondo effect and the concomitant non-Fermi liquid excitations. This raises the prospect that an inherent interplay between non-Fermi liquid and quantum critical physics takes place in other strongly correlated electron systems as well: for high temperature superconductors, such considerations are in line with the scaling behavior observed near the optimal doping.

It is also instructive to relate the quantum critical metal physics with some recent work on quantum critical magnets [15]. The transition considered in the Kondo lattice and that in the quantum magnet are different: in the former it is between a paramagnetic metal and an antiferromagnetic metal and there is a well-defined order parameter only on one side; in the latter the transition considered is between a “valence-bond solid” and an antiferromagnetic insulator, both of which break translational symmetry and each has its distinct order parameter. Nonetheless, there are parallels between the two cases:

- The non-Fermi liquid excitations of a Kondo lattice  $\longleftrightarrow$  the fractionalized spin excitations – “spinons” – of a quantum magnet.
- The Kondo singlets – formed between the local moments and conduction electrons – of a Kondo lattice  $\longleftrightarrow$  the valence bond – or spin singlets formed between the local moments across bonds – of a quantum magnet.

- The Kondo singlet formation in the ground state implies conventional Kondo resonances – spin- $\frac{1}{2}$  and charge- $e$  Landau quasiparticles – in the excitation spectrum of a Kondo lattice  $\longleftrightarrow$  the valence bond formation in the ground state is symptomatic of the confinement of spinons in a quantum magnet.
- The destruction of the Kondo singlet is responsible for the emergence of the non-Fermi liquid excitations at the QCP of a Kondo lattice  $\longleftrightarrow$  the destruction of the valence bond is responsible for the emergence of spinons that are not confined at the QCP of a quantum magnet.
- The non-Fermi liquid excitations and the spinons in a Kondo lattice and a quantum magnet, respectively, are a part of their respective quantum-critical spectrum.
- In both cases, the excitations of the two phases that the QCP separates are conventional.

These parallels arise in spite of the differences in the model and in the underlying physics. They suggest the exciting possibility that a breakdown of the fluctuating order parameter theory occurs over a wide range of strongly correlated systems.

## 6 Summary

We have described some recent developments in the area of quantum critical metals, with an emphasis on the interplay between non-Fermi liquid physics and quantum critical behavior. In addition to the traditional consideration that quantum criticality leads to non-Fermi liquid behavior, the new realization is that the non-Fermi liquid excitations should be treated as part of the quantum-critical theory thereby leading to new classes of quantum phase transitions.

In the specific case of quantum critical heavy fermions, the non-Fermi liquid physics is characterized by a destruction of the Kondo effect. This results in the picture of local quantum criticality. Experimental evidence for this picture has come from inelastic neutron scattering, NMR, Grüneisen ratio, and Hall effect.

Finally, general considerations and specific examples suggest that the breakdown of the order parameter fluctuation theory and the emergence of novel excitations at a quantum critical point are properties relevant to a broad range of strongly correlated matters.

This article is based on a talk given at the DPG Spring Meeting at Regensburg Germany in March 2004. I am grateful to D. Grempel, K. Ingersent, S. Kirchner, E. Pivoarov, S. Rabello, J. L. Smith, J.-X. Zhu, and L. Zhu, as well as the Dresden group (particularly P. Gegenwart, R. Küchler, J. A. Mydosh, S. Paschen, and F. Steglich), P. Coleman, M. Garst, and A. Rosch for collaborations, many colleagues for discussions, and NSF Grant No. DMR-0424125 and the Robert A. Welch foundation for support.

## References

1. For reviews, see K. Wilson and J. Kogut, Phys. Rep. **C12**, 75 (1974); S.-K. Ma, *Modern Theory of Critical Phenomena* (Addison-Wesley, Redwood, 1976).
2. J. Hertz, Phys. Rev. **B14**, 1164 (1976).
3. For a review, see S. Sachdev, *Quantum Phase Transitions* (Cambridge University Press, Cambridge, 1999).
4. Q. Si, S. Rabello, K. Ingersent, and J. L. Smith, Nature **413**, 804 (2001).
5. P. Coleman and C. Pépin, Acta Physica Polonica **B34**, 691 (2003).
6. G. R. Stewart, Rev. Mod. Phys. **73**, 797 (2001).
7. Q. Si, APCTP Bulletin **11-12**, 7-12 (2003); cond-mat/0302110.
8. P. Pfeuty, Ann. Phys. NY **57**, 79-90 (1970).
9. A. J. Millis, Phys. Rev. **B48**, 7183 (1993).
10. N. F. Berk and J. R. Schrieffer, Phys. Rev. Lett. **17**, 433 (1966).
11. S. Doniach and S. Engelsberg, Phys. Rev. Lett. **17**, 750 (1966).
12. T. Moriya, *Spin Fluctuations in Itinerant Electron Magnetism* (Springer, Berlin, 1985).
13. G. G. Lonzarich, in *Electron*, edited by M. Springford (Cambridge Univ. Press, Cambridge, 1997), p. 109.
14. E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
15. T. Senthil, A. Vishwanath, L. Balents, S. Sachdev, and M. P. A. Fisher, Science **303**, 1490 (2004).
16. H. v. Löhneysen *et al.*, Phys. Rev. Lett. **72**, 3262 (1994).
17. J. Custers *et al.*, Nature **424**, 524 (2003).
18. N. D. Mathur *et al.*, Nature **394**, 39 (1998).
19. A. Demuer *et al.*, J. Phys. Cond. Matt. **13**, 9335 (2001).
20. A. Schröder *et al.*, Phys. Rev. Lett. **80**, 5623 (1998).
21. O. Stockert *et al.*, Phys. Rev. Lett. **80**, 5627 (1998).
22. A. Schröder *et al.*, Nature **407**, 351-355 (2000).
23. D. Belitz, T. Kirkpatrick, and T. Vojta, Phys. Rev. B **55**, 9452 (1997).
24. Ar. Abanov and A. V. Chubukov, Phys. Rev. Lett. **84**, 5608 (2000).
25. S. Doniach, Physica **B91**, 231 (1977).
26. C. M. Varma, Rev. Mod. Phys. **48**, 219 (1976).
27. A. C. Hewson, *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, Cambridge, 1993).
28. Q. Si, J. L. Smith and K. Ingersent, Int. Journ. Mod. Phys. **B13**, 2331 (1999); cond-mat/9905006.
29. Q. Si, S. Rabello, K. Ingersent, and J. L. Smith, Phys. Rev. **B68**, 115103 (2003).
30. J.-X. Zhu, D. R. Grempel, and Q. Si, Phys. Rev. Lett. **91**, 156404 (2003).
31. D. Grempel and Q. Si, Phys. Rev. Lett. **91**, 026401 (2003).
32. P. Coleman, C. Pépin and A. M. Tsvelik, Phys. Rev. B **62**, 3852 (2000).
33. T. Senthil, S. Sachdev, and M. Vojta, Phys. Rev. Lett. **90**, 216403 (2003).
34. Q. Si, J. Phys.: Condens. Matter **15**, S2207 (2003).
35. A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
36. W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989).
37. J. L. Smith and Q. Si, Phys. Rev. **B61**, 5184 (2000); Q. Si and J. L. Smith, Phys. Rev. Lett. **77**, 3391 (1996).
38. R. Chitra and G. Kotliar, Phys. Rev. Lett. **84**, 3678 (2000).

39. L. Zhu and Q. Si, Phys. Rev. B**66**, 024426 (2002); G. Zar  nd and E. Demler, Phys. Rev. B **66**, 024427 (2002).
40. P. Sun and G. Kotliar, Phys. Rev. Lett. **91**, 037209 (2003)
41. S. Pankov *et al.*, Phys. Rev. B**69**, 054426 (2004).
42. M. C. Aronson *et al.*, Phys. Rev. Lett. **75**, 725 (1995).
43. D. E. MacLaughlin *et al.*, Phys. Rev. Lett. **87**, 066402 (2001).
44. M. B. Maple *et al.* J. Low Temp. Phys. **95**, 225 (1994).
45. P. Gegenwart *et al.*, Acta Physica Polonica B**34**, 323 (2003).
46. J. D. Thompson, private communications.
47. S. Nakatsuji, private communications.
48. K. Ishida *et al.*, Phys. Rev. B**68**, 184401 (2003).
49. R. E. Walstedt, H. Kojima, N. Butch, and N. Bernhoeft, Phys. Rev. Lett. **90**, 067601 (2003)
50. A. Rosch, Phys. Rev. Lett. **82**, 4280 (1999).
51. L. Zhu, M. Garst, A. Rosch, and Q. Si, Phys. Rev. Lett. **91**, 066404 (2003).
52. R. K  chler *et al.*, Phys. Rev. Lett. **91**, 066405 (2003).
53. S. Paschen *et al.*, preprint (2004).

# Quantum Monte Carlo Simulation of Confined Bosonic Atoms in Optical Lattices

Stefan Wessel<sup>1</sup>, Fabien Alet<sup>1,2</sup>, Matthias Troyer<sup>1,2</sup>, and George Batrouni<sup>3,4</sup>

<sup>1</sup> Theoretische Physik, ETH Zürich  
8093 Zürich, Switzerland

<sup>2</sup> Computational Laboratory, ETH Zürich  
8092 Zürich, Switzerland  
[troyer@phys.ethz.ch](mailto:troyer@phys.ethz.ch)

<sup>3</sup> Institut Non-Linéaire de Nice, Université de Nice-Sophia Antipolis  
1361 route des Lucioles, 06560 Valbonne, France\*

<sup>4</sup> Department of Physics, NTNU  
7491 Trondheim, Norway

**Abstract.** We study static properties of ultra-cold bosonic atoms in two-dimensional optical lattices by quantum Monte Carlo simulations of the bosonic Hubbard model in parabolic confinement potentials. Our focus is on local probes identifying Mott-insulating and superfluid regions, which can coexist in the inhomogenous environment of the trap. By proposing an effective ladder model for the boundary region between the two phases we can show clear evidence for the absence of true quantum critical behavior and explain the absence of critical slowing down at the quantum phase transition in a harmonic trap.

## 1 Introduction

Experiments on ultra-cold atomic gases in optical lattices [1,2] provide the unique opportunity to compare directly models for strongly correlated many body quantum lattice models with near-perfect realizations of those models in experiments. Unlike in experiments on strongly correlated materials, where often the models simulated numerically are prototype *toy models* of the effective models describing the low-temperature properties of materials, the same models are *realistic models* of trapped cold atoms and allow quantitative comparisons.

Experiments performed so far have focused on bosonic atoms [1,2] which are easier to cool than fermions. A quantitative understanding of these bosonic systems will, in the future, allow to interpret correctly experimental measurements on fermionic atoms, which could be used as analog quantum simulators for strongly correlated fermionic systems [3].

Here we will present preliminary results of an extensive set of simulations [4] performed on one-dimensional (1D), two-dimensional (2D) and three-dimensional (3D) systems, expanding on previous work on 1D [5] and 3D [6]

---

\* Permanent address.

systems. The main results will be the identification of a local probe of the state at a given location in the inhomogenous trap, a discussion of the validity of local-potential approximations, and evidence for the absence of quantum criticality in these systems.

## 2 Model and Numerical Techniques

Cold confined atomic Bose gases in an optical lattice are well described by the inhomogeneous Bose Hubbard Hamiltonian:

$$H = -t \sum_{\langle i,j \rangle} (b_i^\dagger b_j + h.c.) + \frac{U}{2} \sum_i n_i(n_i - 1) + V \sum_i r_i^2 n_i - \mu \sum_i n_i, \quad (1)$$

where  $b_i^\dagger$ ,  $(b_i)$  denotes the creation (destruction) operator for bosons at site  $i$ , located a distance  $r_i$  from the center of the trap, and  $n_i = b_i^\dagger b_i$  the local density operator. The nearest-neighbor hopping integral  $t$ , the onsite Hubbard repulsion  $U$ , and the curvature of the harmonic trapping potential  $V > 0$  are tunable parameters in experiments.

While experiments are, to a good approximation, performed at constant particle number, we perform our simulations at constant chemical potential  $\mu$ . Comparisons to simulations with fixed particle numbers indicate that the systems studied here are large enough for the difference in ensemble to be negligible. For the discussions below we introduce a local chemical potential

$$\mu_i^{eff} = \mu - Vr_i^2 \quad (2)$$

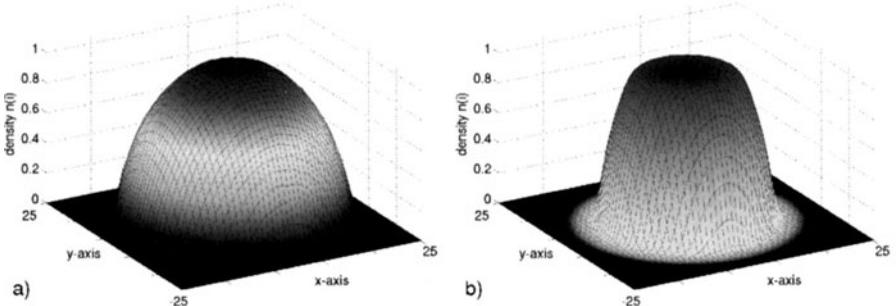
which decreases upon moving away from the trap-center.

Quantum Monte Carlo (QMC) simulations of the Hamiltonian Eq. (1) were performed using the Stochastic Series Expansion method [7,8], with directed loop updates [9,10,11]. This algorithm requires a cutoff  $N_{max}$  on the local site occupation. Performing simulations at mean densities  $\langle n_i \rangle \leq 1$ , a cutoff  $N_{max} = 2$  or  $3$  can be chosen without introducing significant errors. The temperature used in the QMC calculations was chosen low enough to be essentially in the ground-state.

We specify all lengths in units of the lattice constant  $a$  of the optical lattice. In this work, we focus on the 2D case on square lattices with size up to  $50 \times 50$ . This is comparable to the size of the experimentally studied systems, and with the values chosen for the couplings the density outside the boundary is negligible because of the trapping potential.

To distinguish the local phases in the inhomogeneous system and to probe for quantum criticality we define the *local compressibility* at site  $i$ ,

$$\kappa_i^{local} = \frac{\partial N}{\partial \mu_i^{eff}} = \frac{\partial n_i}{\partial \mu} = \beta (\langle n_i n \rangle - \langle n_i \rangle \langle n \rangle), \quad (3)$$



**Fig. 1.** Spatial dependence of local density for a 2D trapped system: a) in the superfluid phase; b) at stronger lattice depths a central Mott-insulating region is formed. The simulations were performed for a harmonic trap with curvature  $V/U = 0.002$ , at  $\mu/U = 0.37$  and for  $U/t = 6.7$  (a),  $U/t = 25$  (b)

by the density response of the total system  $n$  to a local chemical potential change at site  $i$ . This is the same as the response of the local density  $n_i$  to a change in global chemical potential  $\mu$ .

One may also define a compressibility in terms of the variance, as was done in Ref. [5]

$$\kappa_i^{\text{onsite}} = (\langle n_i^2 \rangle - \langle n_i \rangle^2), \quad (4)$$

which we denote as the *onsite compressibility* in the following.

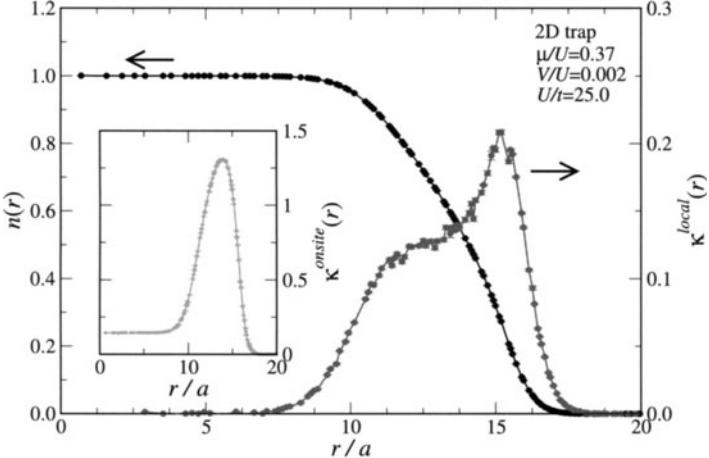
### 3 Simulations of Realistic 2D Traps

#### 3.1 Phase Coexistence in Trapped Systems

Trapped inhomogeneous systems can show coexistence of both superfluid and Mott insulating phases, as has been clearly identified in experiments [1,2], in mean-field investigations [12] and in numerical simulations in 1D [5] and 3D [6]. Simulating large realistic systems in 2D, we can clearly observe this coexistence. At large hoppings  $t$  the total system is superfluid (Fig. 1a).

As the hopping amplitude  $t$  is decreased, a Mott-insulating plateau with integer density (here  $\langle n_i \rangle = 1$ ) forms in the center of the trap, as seen in Fig. 1b. A superfluid ring-like region with a non-uniform particle density surrounds the central plateau. The typical width of this superfluid ring is 6–8 times the lattice unit, for the parameters used in our simulations. Far away from the center, the density is  $\langle n_i \rangle = 0$  which can also be viewed as a Mott-plateau.

The emergence of the Mott plateau and the shrinking of the superfluid region has been interpreted as a quantum phase transition [1]. We prefer to view it as a crossover where the volume fraction of the Mott insulating



**Fig. 2.** Radial dependence of the density  $n$ , and the local compressibility,  $\kappa^{local}$ , within a two-dimensional harmonic trap with curvature  $V/U = 0.002$ , for  $\mu/U = 0.37$  and  $U/t = 25$ , with a superfluid ring surrounding the  $n = 1$  central Mott-plateau. The inset shows the behavior of the onsite compressibility,  $\kappa^{onsite}$ , which remains finite within the Mott-plateau

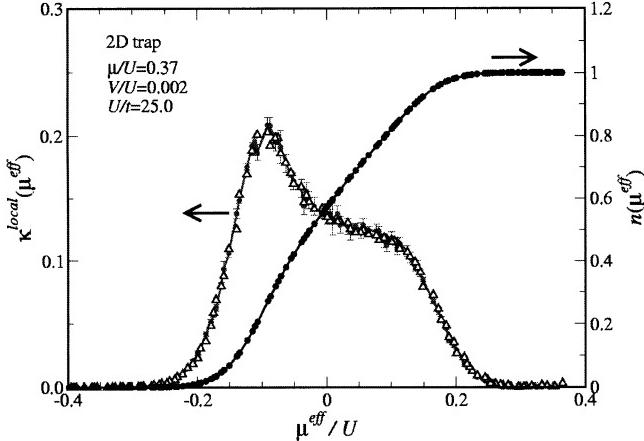
phase grows and that of the superfluid phase shrinks. After discussing the quantitative properties of the two phases and of the boundary region we will, in Sec. 4, show that no quantum critical behavior is observed in this system, which supports our interpretation of this phenomenon as a crossover instead of a phase transition.

### 3.2 Local Compressibility

Figure 1 is useful for a qualitative view, but we need a quantitative probe to distinguish the different phases. This probe is given by the compressibility. In Fig. 2, we show the radial dependence of both the local density  $n(r)$  and the local compressibility  $\kappa^{local}(r)$ , as a function of the distance  $r$  from the center of the trap for the system of Fig. 1b. Again, we observe two well defined Mott regions with local density  $n(r) = 1$  and  $n(r) = 0$ . In these Mott plateaux, the local compressibility vanishes, whereas it is finite in the intervening superfluid ring with non-integral density  $0 < n(r) < 1$ .

More precisely, we observe two well-defined peaks in  $\kappa^{local}(r)$ , signaling an increase in the particle number fluctuations, at the boundary to the Mott regions. While these two peaks are of the same height in hard-core boson models (due to particle-hole symmetry), they are asymmetric in the soft-core model.

The inset of Fig. 2 shows the onsite compressibility  $\kappa^{onsite}(r)$ , which was used in simulations of a 1D trapped system [5]. Contrary to  $\kappa^{local}(r)$ , this quantity does not vanish in the Mott plateau, and displays a pronounced peak



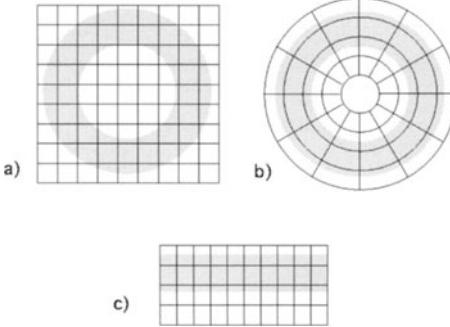
**Fig. 3.** Density,  $n$ , and local compressibility,  $\kappa^{local}$ , for a two-dimensional harmonic trap with curvature  $V/U = 0.002$ , at  $\mu/U = 0.37$  and  $U/t = 25$ , as a function of  $\mu^{eff}$ . The results for  $\kappa^{local}$  from the density fluctuations (Eq. 3) (circles) agree with the numerical derivative  $\partial n/\partial\mu^{eff}$  (triangles)

approximately in the middle of the superfluid ring. These results demonstrate that the local compressibility  $\kappa^{local}(r)$  is the better probe for the existence of superfluid or Mott regions in the system than the onsite compressibility  $\kappa^{onsite}(r)$ . Moreover, the response of the *total* system to a local excitation (for example by changing the laser intensity at one specific point) should be easier to study experimentally than the local response to a local excitation.

### 3.3 Local Potential Approximation

In Fig. 3, we again show the behavior of the local density and compressibility (open circles), but this time as a function of the effective chemical potential  $\mu^{eff}$  as defined in Eq. (2). The observed data collapse of the local density and compressibility and the smooth behavior of these curves indicate that, for the realistic parameters used in the simulations, a local potential approximation holds: the local density can be determined from the value of the local effective chemical potential. The validity of this approximation is further confirmed by the observation that the local compressibility coincides perfectly with the numerical derivative of the curve  $n(\mu^{eff})$ .

While  $n(\mu^{eff})$  seems to be a universal function for this trap geometry it is important to note that, around the boundary it is not the same as  $n(\mu)$  for the homogeneous 2D system without a trap, as we discuss in more detail in Ref. [4]. While a local potential approximation with an effective local chemical potential holds, the value  $n(\mu^{eff})$  has to be taken from simulations in a trap and not the homogeneous system. A key difference is that at the phase transition the density  $n(\mu)$  of the homogeneous 2D system shows a cusp-like behavior due to quantum criticality, which is absent in the trapped system.



**Fig. 4.** Different lattice topologies and expected superfluid regions (denoted in grey) for a quadratic confining potential for such topologies : (a) square lattice, (b) polar lattice, (c) ladder lattice

## 4 Effective Ladder Model for the Superfluid Layers

### 4.1 The Effective Ladder Model

A parabolic potential imposed on a regular hypercubic lattice makes for an irregular form of the superfluid shell surrounding the Mott insulator (see Fig. 4a). In this section we investigate whether this randomness is the reason for the apparent absence of quantum critical behavior by looking at a “smoother” lattice. Although it cannot be realized in experiments, we consider a polar lattice, such as depicted in Fig. 4b, which is topologically equivalent to the ladder system in Fig. 4c. The rotation invariance of the polar lattice (translation invariance of the ladders along the leg direction) simplifies the detection of critical behavior by reducing finite size effects.

While the polar and ladder lattices are not topologically equivalent to the square lattice, universality of quantum critical behavior ensures that we study the same effects. Comparing the lattices will be a further check for the validity of a local potential approximation for the trapped system.

The Hamiltonian of the ladder system of size  $L_x \times L_y$  is

$$H = -t \sum_{i=1}^{L_x} \sum_{j=1}^{L_y} \left( b_{i,j}^\dagger b_{i+1,j} + b_{i,j}^\dagger b_{i,j+1} + h.c. \right) + \frac{U}{2} \sum_{i=1}^{L_x} \sum_{j=1}^{L_y} n_{i,j} (n_{i,j} - 1) - \sum_{j=1}^{L_y} \mu(j) \sum_{i=1}^{L_x} n_{i,j} \quad (5)$$

where the trapping potential of the experiments is now modeled by a chemical potential  $\mu(j)$  which is constant for each leg  $j$  of the ladder. The boundary conditions are chosen to be periodic along the legs of the ladder, and open in the transverse direction.

With these choices, the local density  $\langle n_{i,j} \rangle$  is independent of  $i$  due to translation invariance along a leg of the ladder, and we denote by  $n(j)$  the density of particles on the  $j$ -th leg. Similarly the local compressibility  $\kappa^{local}(j)$  is the same for all sites on the  $j$ -th leg. The results shown below have been obtained for a geometry with  $L_x = 64$ , and  $L_y = 10$ .

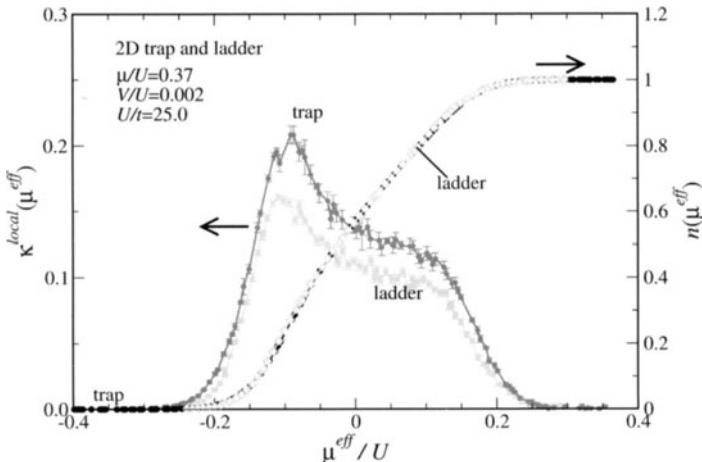
To simplify simulations, we linearize the quadratic potential  $\mu + Vj^2$ ,

$$\mu(j) = \mu_0 + j\Delta\mu, \quad (6)$$

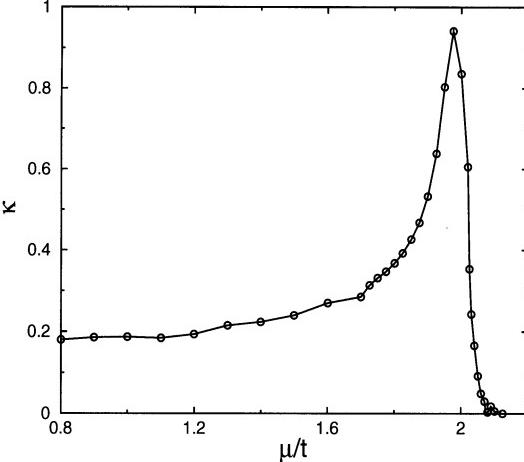
where  $\Delta\mu$  is the difference in chemical potential between two neighboring legs. In our simulations we fix  $\Delta\mu/U = 0.053$  and make sure that the first leg is always in the  $\langle n \rangle = 0$  Mott insulating phase and the last leg in the  $\langle n \rangle = 1$  insulating phase. Sweeping the value of the global  $\mu_0$  allows us to obtain results for all values of the local chemical potentials, and to drive one of the legs from the superfluid to the Mott insulating phase.

## 4.2 Results for the Ladder Model Compared with a Realistic Trap

In Fig. 5 we show the combined results of all simulations with different values of  $\mu_0$ . We plot both the local density  $\langle n(j) \rangle$  and local compressibility  $\kappa^{local}(j)$  for all legs and all values of  $\mu_0$  as a function of the local chemical potential. Notice that all data from simulations using different values of  $\mu_0$  collapse onto a single curve. Due to the equivalence of all sites on the ladder with the same value of the chemical potential, this data collapse is expected in the ladder model.



**Fig. 5.** Density,  $n$ , and local compressibility,  $\kappa^{local}$ , for a two-dimensional harmonic trap on a square lattice and for a ladder lattice, as a function of  $\mu^{eff}$ . The parameters for the trapping potential are  $V/U = 0.002$ , at  $\mu/U = 0.37$ , and the parameters of the ladder model are chosen in such a way that they cover the whole superfluid region (see text). The density curves for the trapped square and for the ladder lattices coincide almost perfectly. The small differences between them are manifested by the two different compressibility curves, which have the same overall shape



**Fig. 6.** Compressibility  $\kappa$  versus chemical potential for a single 1d chain of hardcore bosons, of length  $L_x = 64$ . The compressibility shows signature of a square root divergence at the transition (the finite size of the system studied forbids a true divergence)

Fig. 5 also shows the results obtained for the realistic trap simulations, as described in the previous section. The results *coincide almost perfectly*, which is a strong justification for the validity of the effective model and the applicability of a local potential approximation.

The small differences can be attributed to the different shape of the trapping potential which is parabolic for the realistic trap, but linearized in the ladder model. While the differences in the densities are small, the compressibility, which is the derivative of the densities, is more sensitive.

### 4.3 Absence of Quantum Critical Behavior

Having shown that the ladder model captures the important features of the realistic trap even quantitatively we now investigate the possibility of quantum critical behavior at the phase transition between superfluid and Mott insulator, by comparing to a homogeneous model.

For comparison we show, in Fig. 6, the compressibility of a hardcore boson model on a chain of the same size  $L_x = 64$ . Already on such a short chain a sharp peak is visible, which diverges as the chain length  $L_x$  is increased. One could expect to see similar quantum critical behavior as we change the chemical potential  $\mu_0$  in the ladder model to drive one of the legs across the phase transition.

Our results clearly show that this is not the case: while the broad peaks in Fig. 5 are remnants of this quantum critical behavior, they do not diverge as the system size in the ladder model is increased. The effective ladder model and also the realistic trapped system *do not show quantum critical*

*behavior* if the system size is increased while keeping the gradient of the trapping potential constant. The quantum critical behavior that might have been expected to occur in the boundary layer between the Mott insulator and the superfluid is destroyed by the coupling to the rest of the system. More detailed results and discussions will be presented in Ref. [4].

## 5 Discussion and Conclusion

Our QMC simulations provide insight into the physics of trapped bosonic systems on lattices. The validity of a local potential approximation, where local quantities such as the local density or compressibility depend mainly on the value of the local chemical potential, is confirmed by an excellent data collapse of local quantities on single curves. This single curve is however *not* the same as for the homogeneous bulk system; the differences are particularly pronounced in the interesting vicinity of the transition between Mott insulator and superfluid. There the singularities due to quantum critical behavior are removed and replaced by smooth and broad features.

While the bulk behavior can give qualitative insight into the behavior of trapped systems, quantitative results can only be obtained by QMC simulations of realistic systems. Results for realistic 2D traps of a size comparable to experiments have been presented here, and more details, in particular on anisotropic systems will be presented in a forthcoming publication [4]. Three-dimensional simulations have so far been performed only on systems with linear dimensions 2-3 times smaller than experimental realizations, but realistic simulations will be possible in the near future using faster computers and improved algorithms.

An effective ladder model, which quantitatively models the realistic trapped system, provides evidence for the absence of quantum critical behavior. The ladder model allows us to exclude the randomness imposed on the superfluid ring due to the underlying square lattice as the source of this absence of quantum criticality. Instead, the divergences due to quantum critical fluctuations are suppressed and rounded by the coupling to the rest of the system. It will be very interesting to develop an effective action for this coupling, which might explain the power law behavior observed in Ref. [13].

The absence of quantum critical behavior in traps of sizes comparable to experiments also agrees well with the fast dynamics of the “phase transition” and the observed absence of “critical slowing down” of the dynamics in experiments [1]. Critical slowing down at a second order phase transition is caused by the long time scales it takes to establish a uniform order parameter across the whole system. Small ordered domains, with differently broken  $U(1)$  symmetry are quickly formed, but as these domains grow and merge, the dynamics to establish the same  $U(1)$  symmetry breaking across neighboring domains slows down as the domains grow in size. The dynamics of the “quantum phase transition” in the trapped system is different: driving

the system from the superfluid to the Mott insulator happens by nucleating a small Mott domain, which then grows as the depth of the optical lattice is increased. As the Mott phase grows in volume and the superfluid phase shrinks the “quantum phase transition” is observed. This is however better viewed as a crossover with changing volume fractions of the two phases than as a phase transition: the large Mott plateau is always surrounded by a shell of coherent superfluid. Sweeping back to the superfluid phase by decreasing the depth of the optical lattice, the Mott insulator melts and atoms join the superfluid. The dynamics here is not that of two large domains merging, but that of a single atom joining the coherent superfluid and there is *no critical slowing down* involved in this process. It might be possible to experimentally observe critical slowing down in a trap by first driving the system deep into the Mott insulating region, then kicking it to destroy the phase coherence in the remaining superfluid shell and afterwards quickly driving it back into the superfluid.

### Acknowledgements

The numerical calculations were done using the SSE application package with generalized directed loop techniques [11] of the ALPS project [14], and performed on the Asgard Beowulf cluster at ETH Zürich. We thank A. Muramatsu, M. Rigol, and N. V. Prokof’ev for fruitful discussions. S.W., F.A. and M.T. acknowledge support from the Swiss National Science Foundation. G.B. is supported by the NSF-CNRS cooperative grant #12929 and thanks the NTNU, the Complex Group and Norsk Hydro for their hospitality and generosity during a sabbatical stay.

### References

1. M. Greiner *et al.*, Nature **415**, 39 (2002).
2. T. Stöferle *et al.*, cond-mat/0312440.
3. W. Hofstetter, J. I. Cirac, P. Zoller, E. Demler, and M. D. Lukin, Phys. Rev. Lett. **89**, 220407 (2002).
4. S. Wessel, F. Alet, M. Troyer and G. Batrouni, wnd-mat/0404552.
5. G. G. Batrouni *et al.*, Phys. Rev. Lett. **89**, 117203 (2002).
6. V. A. Kashurnikov, N. V. Prokof’ev, and B. V. Svistunov, Phys. Rev. A **66** 031601(R) (2002).
7. A.W. Sandvik and J. Kurkijärvi, Phys. Rev. B **43**, 5950 (1991); A. W. Sandvik, J. Phys. A **25**, 3667 (1992).
8. A.W. Sandvik, Phys. Rev. B **59**, R14157 (1999).
9. O. F. Syljuåsen and A. W. Sandvik, Phys. Rev. E **66**, 046701 (2002).
10. O. F. Syljuåsen, Phys. Rev. E **67**, 046701 (2003).
11. F. Alet, S. Wessel and M. Troyer, cond-mat/0308495.
12. D. Jaksch, C. Bruder, J. I. Cirac, C. W. Gardiner, and P. Zoller, Phys. Rev. Lett. **81**, 3108 (1998).

13. M. Rigol, A. Muramatsu, G.G. Batrouni, and R.T. Scalettar Phys. Rev. Lett. **91**, 130403 (2003).
14. M. Troyer *et al.*, Lecture Notes in Computer Science **1505**, 191 (1998). Source codes of the libraries can be obtained from <http://alps.comp-phys.org/>.

# Cold Atoms in Optical Lattices

Wilhelm Zwerger

Institut für Theoretische Physik, Universität Innsbruck  
Technikerstr. 25, 6020 Innsbruck, Austria

**Abstract.** Cold atoms in optical lattices provide a novel class of many body systems with widely tunable interactions. They allow quantum state engineering of complex many body states and a study of their associated quantum phase transitions, including dynamics. Basic phenomena in strong correlation physics like the Mott-Hubbard-transition, the BCS-BEC-crossover, Luttinger-liquids or quantum impurity problems can be realized. As two simple examples, we discuss the superfluid to Mott-insulator transition of Bosons and Luttinger-liquid physics in 1D atomic wires.

## 1 Introduction

The realization of Bose-Einstein condensation (BEC) in dilute atomic gases has opened a wide area of research in atomic physics, where quantum-statistical effects are of crucial importance: upon cooling, bosonic atoms in a trap condense into a superfluid state at a rather sharply defined critical temperature  $T_c$  [1,2]. In turn, cold fermions continuously evolve into a degenerate gas, resisting spatial compression due to their Fermi pressure [3]. Both features are a consequence of the purely statistical interaction between the atoms. By contrast, the actual interparticle potential plays only a comparatively minor role which is due – of course – to the extreme diluteness of these gases. For a quantitative argument, we start from the standard pseudopotential description, which replaces the complicated interatomic potential by an effective contact interaction of the form

$$U(\mathbf{x}) = \frac{4\pi\hbar^2 a_s}{m} \cdot \delta(\mathbf{x}) = g \cdot \delta(\mathbf{x}) \quad (1)$$

containing the exact s-wave scattering length  $a_s$  as the only parameter. For identical fermions, there is no s-wave scattering due to the Pauli-principle. To lowest order we thus obtain an ideal Fermi gas, whose many body wavefunction is a Slater-Determinant of the  $N$  lowest single particle eigenstates in a given confining potential. For bosons, in turn,  $a_s$  is finite, however at a given density  $n$  the importance of direct interaction effects can be estimated from the ratio

$$\gamma = \frac{\epsilon_{int}}{\epsilon_{kin}} = \frac{gn}{\hbar^2 n^{2/3}/m} \approx n^{1/3} a_s \quad (2)$$

between the interaction and the kinetic energy per particle. Now the average interparticle spacing  $n^{-1/3}$  is of the order of  $\mu\text{m}$  while the scattering length is usually in the 10 nm range. The ratio  $\gamma$  is therefore very small, with typical values around 0.02. This puts us into the weak coupling limit, where the many body ground state of  $N$  bosons is well approximated by a simple product [2]

$$\Psi_{GP}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \prod_{i=1}^N \phi(\mathbf{x}_i) \quad (3)$$

in which all atoms are in the identical single particle state  $\phi(\mathbf{x})$ . Taking (3) as a variational ansatz, the optimal ‘macroscopic wave function’  $\phi(\mathbf{x})$  is found to obey the well known Gross-Pitaevski equation. It describes – even on a quantitative level – a wealth of remarkable and nontrivial properties of trapped condensates from interference between different condensates to collective modes or vortices. From a many body point of view, the effective single particle or Hartree-description (3) is of course the simplest of all possible cases, containing no interaction induced correlations between different atoms at all. A first step to go beyond this mean field description is the well known Bogoliubov theory. This is usually introduced by considering small fluctuations around the Gross-Pitaevski equation in a systematic expansion in the number of noncondensed particles [4]. From a many body point of view, however, it is more instructive to formulate Bogoliubov theory in such a way that the many body boson ground state is approximated by an optimized product [2]

$$\Psi_{Bog.}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \prod_{i < j} \phi_2(\mathbf{x}_i, \mathbf{x}_j) \quad (4)$$

of identical, symmetric two particle wave functions  $\phi_2$ . This allows us to build in the interaction beyond the mean field potential by suppressing configurations in which particles  $i$  and  $j$  are close together. The many body state thus incorporates two-particle correlations which are important e.g. to obtain the standard sound modes and the related coherent superposition of ‘particle’ and ‘hole’ excitations. However, even the Bogoliubov description is restricted to the regime  $\gamma \ll 1$ , where interactions lead only to a small depletion of the condensate at zero temperature. The associated ground state is again characterized by a macroscopic matter wave field and continuously evolves from that of a noninteracting gas.

An obvious way to go beyond the weak coupling regime is to increase the dimensionless interaction strength parameter  $\gamma$  by changing the scattering length via a Feshbach resonance [5]. Such a resonance occurs, when a bound state in a closed channel crosses zero, which is the energy of two colliding atoms in the s-wave scattering limit. The resonances are accessible e.g. by changing the magnetic field  $B$  which shifts the bound state energies due to

the hyperfine interaction. On a qualitative level, Feshbach resonances are described by a scattering length

$$a_s(B) = a_{bg} \left( 1 - \frac{\Delta B}{B - B_0} \right) \quad (5)$$

which diverges at a specific magnetic field value  $B_0$ , approaching a background scattering length  $a_{bg}$  when  $|B - B_0|$  is much larger than the width  $\Delta B$  of the resonance. For fermions, where the Pauli principle strongly suppresses three-particle collisions, Feshbach resonances are a major tool to manipulate the interactions. Indeed, they have been the crucial ingredient for the very recent realization of molecular condensates in  ${}^6\text{Li}$  and  ${}^{40}\text{K}$  [6,7,8]. For bosons, the use of Feshbach resonances to enter the strong coupling regime is limited by the fact that the associated condensate lifetime strongly decreases due to three-body losses which occur at a rate [9]

$$\dot{n}/n = -\text{const.} \frac{\hbar}{m} (na_s^2)^2. \quad (6)$$

In spite of this problem, Feshbach resonances have been successfully used to enter the strong coupling regime for bosons, see e.g. [10]. In the following, we will discuss an alternative route to reach strong coupling even at small values of  $n^{1/3}a_s$ . It is based on confining cold atoms in the periodic potential of an optical lattice generated via the dipole force  $\mathbf{F} = -\nabla(\alpha(\omega)|\mathbf{E}(\mathbf{x})|^2/2)$  due to the ac Stark shift which atoms experience in a standing, off-resonant light field. Depending on the sign of the polarizability  $\alpha(\omega)$ , the atoms are attracted either to the nodes or the anti-nodes of the laser intensity. In this manner, one-, two-, or three-dimensional lattices can be created with a lattice constant  $a = \lambda/2$  which is half the laser wavelength (typically  $a$  is in the range between  $0.5\ \mu\text{m}$  and  $5\ \mu\text{m}$ ). In the simplest case, three orthogonal, independent standing laser fields with wave vector  $k$  produce a separable 3D lattice potential

$$V(x, y, z) = V_0 (\sin^2 kx + \sin^2 ky + \sin^2 kz) \quad (7)$$

with a tunable amplitude  $V_0$ . A convenient measure for the strength  $V_0$  of the lattice potential is the recoil energy  $E_r = \hbar^2 k^2 / 2m$  which is typically in the few kHz range. In a deep optical lattice with  $V_0 \gg E_r$ , the energy  $\hbar\omega_0 = 2E_r (V_0/E_r)^{1/2}$  of local oscillations in the well is much larger than the recoil energy and each well supports several quasi-bound states. Provided that all the atoms are in the lowest vibrational level at each site, their kinetic energy is frozen except for the small tunneling amplitude to neighbouring sites. The atoms are then effectively confined to move in the lowest band of the lattice. With  $|l\rangle$  as the states localized at site  $l$ , the appropriate single particle eigenstates are Bloch-waves  $|\mathbf{q}\rangle = \sum_l \exp i\mathbf{q} \cdot \mathbf{l} |l\rangle$  with quasimomentum  $\mathbf{q}$  and energy

$$\epsilon(\mathbf{q}) = \frac{3}{2} \hbar\omega_0 - 2J (\cos q_x a + \cos q_y a + \cos q_z a) \quad (8)$$

The bandwidth parameter  $J$  is essentially the gain in kinetic energy due to nearest neighbour tunneling. In the limit  $V_0 \gg E_r$  it can be obtained from the exact result for the width of the lowest band in the 1D Mathieu-equation

$$J = \frac{4}{\sqrt{\pi}} E_r \left( \frac{V_0}{E_r} \right)^{3/4} \exp -2 \left( \frac{V_0}{E_r} \right)^{1/2} \quad (9)$$

Obviously, in a lattice, it is  $J$  which plays the role of the kinetic energy per particle. The effective value of  $\gamma = \epsilon_{int}/\epsilon_{kin}$  is therefore very large in optical lattices, increasing exponentially with  $V_0/E_r$ . Thus it is the quenching of the kinetic energy for motion in the lowest band which drives cold atoms into the strong coupling regime, even though  $n^{1/3}a_s$  may still be much smaller than one. Alternatively, one may argue that  $\epsilon_{kin}$  becomes small because atoms in a deep optical lattice have an exponentially large effective mass  $m^* = \hbar^2/2a^2J$ . To really obtain interesting many body effects it is of course necessary to have the interaction and the kinetic energy of the same order. This requires optical lattices in which the number of atoms per site is of order one or larger, a regime which could be reached only with cold atoms with a sufficiently high density, as realized in a BEC.

## 2 Mott-Hubbard Transition for Bosons

As a generic example illustrating how cold atoms in optical lattices can be used to study genuine many body phenomena in dilute gases, I will discuss the Mott-Hubbard transition for bosonic atoms which was observed in a truly remarkable experiment by Greiner et. al.[11]. The original idea suggesting the possibility of nontrivial many body states using cold atoms in optical lattices is due to Jaksch et.al. [12]. Their starting point is the Bose-Hubbard model, originally introduced by Fisher et.al. [13] as a model to describe the destruction of superfluidity due to strong interactions or disorder. In its simplest version, it describes bosons hopping with amplitude  $J$  to nearest neighbors on a regular lattice of sites  $l$ . The particles interact with a zero-range, on-site repulsion  $U$ , disfavouring configurations with more than one atom at a given site. With  $\hat{b}_l^\dagger$  as the creation operator of a boson at site  $l$  and  $\hat{n}_l = \hat{b}_l^\dagger \hat{b}_l$  the associated number operator, the Hamiltonian reads

$$\hat{H} = -J \sum_{\langle ll' \rangle} \hat{b}_l^\dagger \hat{b}_{l'} + \frac{U}{2} \sum_l \hat{n}_l (\hat{n}_l - 1) + \sum_l \epsilon_l \hat{n}_l \quad (10)$$

Here  $\langle ll' \rangle$  denotes a sum over nearest neighbour pairs, including double counting. The last term with a variable on-site energy  $\epsilon_l$  is introduced to describe the effect of the trapping potential and acts like a spatially varying chemical potential. The form of the interaction term is precisely that obtained by viewing each site as a local condensate with a Gross-Pitaevski mean field

potential. The relevant interaction parameter  $U$  is thus given by an integral over the on-site wave function  $w(\mathbf{x})$  via

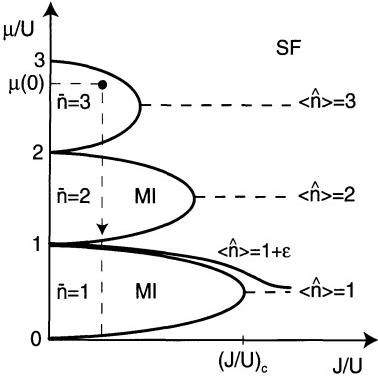
$$U = g \int |w(\mathbf{x})|^4 = \sqrt{\frac{8}{\pi}} k a_s E_r \left( \frac{V_0}{E_r} \right)^{3/4} \quad (11)$$

The explicit result is obtained by taking  $w(\mathbf{x})$  as the Gaussian ground state in the local oscillator potential around any of the sites. More precisely,  $w(\mathbf{x})$  is the exact Wannier wave function of the lowest band. In a separable periodic potential like that in (7),  $w(\mathbf{x})$  decays exponentially in all directions rather than in a Gaussian manner, however this does not seriously affect the calculation of  $U$  in the deep lattice limit  $V_0 \gg E_r$ . As equation (11) shows, the strength of the repulsion increases with  $V_0$ , which is a simple consequence of the squeezing of the on-site wave function  $w(\mathbf{x})$ . With increasing  $V_0/E_r$ , therefore, not only does the kinetic energy drop exponentially, but at the same time the interaction energy increases. As a result, it is possible to reach the strong coupling regime  $U \approx J$  simply by increasing the depth of the optical lattice potential. Regarding the requirements necessary for the validity of the discrete lattice model (10), it is obvious that the atoms have to remain in the lowest vibrational state at each site even in the presence of strong interactions. We thus require that  $U \ll \hbar\omega_0$  which is well obeyed even in deep optical lattices as long as  $ka_s \ll 1$ .

The zero temperature phase diagram of the homogeneous Bose-Hubbard model was first discussed by Fisher et.al. [13]. Although there are considerable quantitative differences between the case of one-, two- or three-dimensional lattices, the qualitative structure is similar in all cases and is shown schematically in Figure 1. At large  $J/U$  the kinetic energy dominates and the ground state is a delocalized superfluid (SF). At small values of  $J/U$ , interactions dominate and one obtains a series of so called Mott-insulating (MI) phases with fixed integer filling  $\bar{n} = 1, 2, \dots$  depending on the value of the chemical potential  $\mu$ . To understand the peculiar structure of these ‘Mott-lobes’, consider first the case of unit filling, i.e. the number  $N$  of atoms is precisely equal to the number  $M$  of lattice sites (in the experiment [11], the total number of lattice sites was about  $1.5 \cdot 10^5$ ). In the limit where  $V_0$  is very large compared to  $E_r$ , hopping of the atoms is negligible and the obvious ground state

$$|\Psi_{MI}\rangle(J=0, \bar{n}) = \prod_1 (|\bar{n}\rangle_1) \quad (12)$$

is a simple product of local Fock-states with precisely one atom ( $\bar{n} = 1$ ) per site. Upon lowering  $V_0$ , the atoms start to hop around, which necessarily involves double occupancy, increasing the energy by  $U$ . Now as long as the gain  $J$  in kinetic energy due to hopping is smaller than  $U$ , the atoms remain localized in the sense that their mean square displacement is finite. For any  $J \neq 0$ , however, the ground state is no longer a simple product state as in (12). Once  $J$  becomes of order or larger than  $U$ , the gain in kinetic energy



**Fig. 1.** Schematic zero temperature phase diagram of the Bose-Hubbard model. The dashed lines of constant integer density  $\langle \hat{n} \rangle = 1, 2, 3$  in the SF hit the corresponding MI phases at the tips of the lobes at a critical value of  $J/U$ , which decreases with density  $\bar{n}$ . For  $\langle \hat{n} \rangle = 1 + \varepsilon$  the line of constant density stays outside the  $\bar{n} = 1$  MI because a fraction  $\varepsilon$  of the particles remains superfluid down to the lowest values of  $J$ . In an external trap with a  $\bar{n} = 3$  MI phase in the center, a series of MI and SF regions appear by going towards the edge of the cloud, where the local chemical potential has dropped to zero

outweighs the repulsion due to double occupancies and the atoms will be delocalized over the whole lattice. In the limit  $J \gg U$ , the many body ground state becomes simply an ideal BEC where all  $N$  atoms are in the  $\mathbf{q} = 0$  Bloch-state of the lowest band. Including the normalization factor in a lattice with  $M$  sites, this state can be written in the form

$$|\Psi_{SF,N}\rangle(U=0) = \left( \frac{1}{\sqrt{M}} \sum_{l=1}^M \hat{b}_l^\dagger \right)^N |0\rangle. \quad (13)$$

For large enough  $J$  therefore, we recover a Gross-Pitaevski like description in terms of one, macroscopically occupied state. In two- and three-dimensional lattices, the critical value for the transition from a MI to a SF is reasonably well described by a mean-field approximation, giving  $(U/J)_c = 5.8z$  for  $\bar{n} = 1$  and  $(U/J)_c = 4\bar{n}z$  for  $\bar{n} \gg 1$ . Here  $z$  is the number of nearest neighbours. In one dimension there are strong deviations from a mean-field approximation and the corresponding values are  $(U/J)_c = 3.37$  for  $\bar{n} = 1$  [14] and  $(U/J)_c = 2.2\bar{n}$  for  $\bar{n} \gg 1$ . The latter result follows from mapping the Bose-Hubbard model to a chain of Josephson junctions, for which the critical value of the transition to a MI phase is known precisely. Using the expressions (9) and (11) for  $J$  and  $U$  in terms of the optical lattice parameters, the critical value of the dimensionless lattice depth  $V_0/E_r|_c$  is obtained from

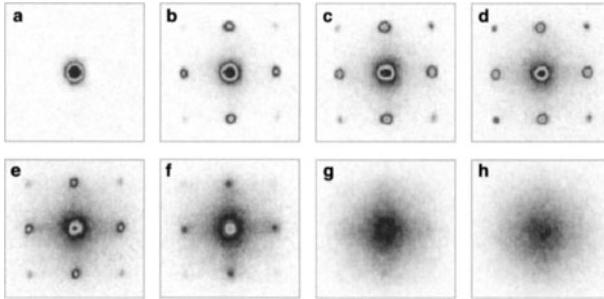
$$\frac{a_s}{a} \cdot \exp 2(V_0/E_r)^{1/2} = \frac{\sqrt{2}}{\pi} \cdot (U/J)_c \quad (14)$$

As an example, consider a simple cubic lattice with  $z = 6$  and use the mean field result for  $(U/J)_c$ . The experimental parameters  $a = 426$  nm and  $a_s = 5.7$  nm [11] then give a critical value  $V_0/E_r|_c = 12.47$  for the SF-MI transition with  $\bar{n} = 1$ . This value is easily accessible experimentally and is also in a regime, where the Bose-Hubbard model with its truncation to the lowest vibrational state in each well indeed applies.

Consider now a filling with  $\langle \hat{n} \rangle = 1 + \varepsilon$  which is slightly larger than one. For large  $J/U$  the ground state has all the atoms delocalized over the whole lattice and the situation is hardly different from the case of unit filling. Upon lowering  $J/U$ , however, the line of constant density remains slightly above the  $\bar{n} = 1$  ‘Mott-lobe’, and stays in the SF regime down to the lowest  $J/U$  (see Fig. 1). For any noninteger filling, therefore, the ground state remains SF as long as the atoms can hop at all. This is a consequence of the fact, that even for  $J \ll U$  there is a small fraction  $\varepsilon$  of atoms which remain SF on top of a frozen MI-phase with  $\bar{n} = 1$ . Indeed this fraction can still gain kinetic energy by delocalizing over the whole lattice without being blocked by the repulsive interaction  $U$  because two of those particles will never be at the same place. The same argument applies to holes when  $\varepsilon$  is negative.

In order to describe the inhomogeneous situation in a weak harmonic trap, we use the standard approximation that a slowly varying external potential may be accounted for by a spatially varying chemical potential  $\mu_l = \mu(0) - \epsilon_l$  (we choose  $\epsilon_l = 0$  at the trap center). Assuming e.g. that the chemical potential  $\mu(0)$  at trap center falls into the  $\bar{n} = 3$  ‘Mott-lobe’, one obtains a series of MI domains separated by a SF by moving to the boundary of the trap where  $\mu_l$  vanishes (see Fig. 1). In this manner, all the different phases which exist for given  $J/U$  below  $\mu(0)$  are present simultaneously. Recalling that the defining property of a MI-phase is its incompressibility  $\partial n / \partial \mu = 0$ , while the SF -phase has a finite value of  $\partial n / \partial \mu$ , the two phases in a trap can be distinguished by their local density. In fact, similar to the succession of incompressible and compressible regions (edge states) near the boundary of a quantum Hall fluid, the atomic density decreases from its maximum value in the trap center to zero in a series of steps. It stays constant in the Mott-phases, even though the external trapping potential is rising. An estimate for the width of the incompressible domains is obtained by noting that for  $J \ll U$  the range in chemical potential over which the density remains constant is close to  $U$ . In a quadratic confining potential with axial frequency  $\nu_z \approx 40$  Hz and with typical values  $U/h = 1$  kHz, the width of the incompressible MI-states is around  $10 \mu\text{m}$ . It remains an experimental challenge to spatially resolve the SF and MI phases in a trap, thus verifying the crucial property of incompressibility.

In practice, the observation of the SF to MI transition is done by absorption imaging the atomic cloud after a given expansion time [11]. The corresponding series of images is shown in Fig. 2 for different values of  $V_0$ , ranging between  $V_0 = 0$  (a) and  $V_0 = 20E_r$  (h). One observes a series of



**Fig. 2.** Absorption images of the interference pattern from atoms in an optical lattice with varying values of the potential depth  $V_0/E_r = 0$  (a), 3 (b), 7 (c), 10 (d), 13 (e), 14 (f), 16 (g) and 20 (h). Taken from ref. [11]

'Bragg-peaks' around the characteristic 'zero-momentum' peak of a condensate in the absence of an optical lattice. With increasing  $V_0$  these peaks become more pronounced. Beyond a critical lattice depth around  $V_0 = 13E_r$  (e), which agrees very well with the above estimate for the SF-MI transition for one Boson per site, this trend is suddenly reversed, however, and the 'Bragg peaks' eventually disappear completely. In order to understand why these pictures indeed provide a direct evidence for a SF to MI transition predicted by the Bose-Hubbard model, we neglect the inhomogeneous nature of the atomic cloud and assume that the absorption images simply reflect the momentum distribution as in a standard time of flight experiment. For atoms which are confined to move in the lowest band of the lattice, it is straightforward to show that the momentum distribution

$$n(\mathbf{k}) = n|w(\mathbf{k})|^2 \sum_{\mathbf{R}} e^{i\mathbf{k} \cdot \mathbf{R}} \rho_1(\mathbf{R}) \quad (15)$$

can be expressed in terms of the exact one-particle density matrix  $\rho_1(\mathbf{R}) = \langle \hat{b}_{\mathbf{R}}^\dagger \hat{b}_0 \rangle$  at separation  $\mathbf{R}$  and the Fourier transform  $w(\mathbf{k})$  of the associated Wannier wave-function. The summation in (15) is over all lattice vectors  $\mathbf{R}$ , which are integer multiples of the three primitive vectors of the given lattice. Now the SF and MI phases are distinguished quite generally by the behaviour of the one particle density matrix at large separation. In the SF,  $\rho_1(\mathbf{R})$  approaches a finite value  $\lim_{|\mathbf{R}| \rightarrow \infty} \langle \hat{b}_{\mathbf{R}}^\dagger \hat{b}_0 \rangle = n_0/n$  which defines the condensate density  $n_0$  (note that this holds only in the 2D and 3D case, while in 1D even the ground state exhibits an algebraic decay as shown by equation (23) below). For the MI phase, in turn,  $\rho_1(\mathbf{R})$  decays to zero exponentially. The SF phase of cold atoms in an optical lattice can thus quite generally be characterized by the fact that at reciprocal lattice vectors  $\mathbf{k} = \mathbf{G}$  defined by  $\mathbf{G} \cdot \mathbf{R} = 2\pi$  times an integer, the momentum distribution  $n(\mathbf{k} = \mathbf{G})$  has a peak

$$n(\mathbf{k} = \mathbf{G}) = N \cdot n_0 |w(\mathbf{G})|^2 \quad (16)$$

which scales with the total number  $N$  of particles. This is the expected behaviour for the interference pattern from a periodic array of phase coherent sources of matter waves and is precisely analogous to the more standard Bragg-peaks in the static structure factor of a solid, with the condensate fraction playing the role of the Debye-Waller factor. The fact that the peaks in the momentum distribution at  $\mathbf{k} = \mathbf{G}$  initially grow with increasing depth of the lattice potential is a result of the strong decrease in spatial extent of the Wannier function  $w(\mathbf{x})$ , which entails a corresponding increase in its Fourier transform  $w(\mathbf{k})$  at higher momenta. It is important to realize that there is no broadening of the peaks as long as  $\rho_1(|\mathbf{R}| \rightarrow \infty)$  is finite, in agreement with the experimental observations [11]. In the MI regime, where  $\rho_1(\mathbf{R})$  decays to zero, remnants of the ‘Bragg-peaks’ still remain (see e.g. (f) in Fig. 2) as long as  $\rho_1(\mathbf{R})$  extends over several lattice spacings, because the series in (15) adds up constructively at  $\mathbf{k} = \mathbf{G}$ . Physically this reflects the fact that phase coherence is still present over distances much larger than one lattice spacing provided one is close to the transition to superfluidity. In contrast to the SF regime, however, these peaks are now broadened and do not scale with the total number  $N$  of particles. In the extreme MI limit  $J \ll U$ , hopping of atoms is negligible and  $\rho_1(\mathbf{R})$  vanishes beyond  $\mathbf{R} = 0$ . Coherence is then completely lost and the momentum distribution is a structureless Gaussian, reflecting the Fourier transform of the Wannier wave function (see (h) in Fig. 2). The quantum phase transition from a SF to a Mott-insulator therefore shows up directly in the interference pattern, which – for the homogeneous system – reveals the existence or not of off-diagonal long range order in the one particle density matrix. The relevant order parameter is the condensate fraction. Of course the actual system is not homogeneous and a numerical computation of the interference pattern is necessary for a quantitative comparison with experiment. This has been done for the 3D case [15] and – in more detail – for a 1D Bose-Hubbard model in a harmonic trap [16,17]. Due to the finite size and the fact that different MI phases are involved, the pattern evolves continuously from the SF to the MI regime. Indeed, as is evident from the phase diagram in Fig. 1, the critical value of  $J/U$  is different for the two different MI phases  $\bar{n} = 1$  and  $\bar{n} = 2$  which are present in the experiment [11]. Nevertheless, a rather sharp transition is observed, because – as (14) shows –  $J/U$  depends exponentially on the control parameter  $V_0/E_r$ . The small change from  $V_0 = 13E_r$  in (e) to  $V_0 = 14E_r$  in (f) thus covers a range in  $J/U$  wider than that, which would be required to distinguish the  $\bar{n} = 1$  from the  $\bar{n} = 2$  transition.

A second signature of the SF to MI transition is the appearance of a finite excitation gap  $\Delta \neq 0$  in the MI. Deep in the MI phase, this gap has size  $U$ , which is just the increase in energy if an atom tunnels to an already occupied adjacent site (note that this is much smaller than the gap  $\hbar\omega_0$  for the excitation of the next vibrational state). The existence of this gap has been verified experimentally by applying a phase gradient in the MI and

measuring the resulting excitations produced in the SF at smaller  $V_0/E_r$  [11]. In this manner the fact that  $\Delta(J \ll U) = U$  was verified for a range of  $V_0$ , all reasonably deep in the MI phase. For reasons discussed above, however, it has not been possible to see the vanishing of the gap near the transition, which is expected to scale like  $\Delta \sim (J_c - J)^{1/2}$  in the 3D case [13]. In the SF regime, there is no excitation gap and instead the homogeneous system exhibits a sound like mode with frequency  $\omega(q) = cq$ . The associated sound velocity follows from the thermodynamic relation  $mc^2 = n_s \cdot \partial\mu/\partial n$  and thus gives information about the superfluid density  $n_s$ . The existence of a sound like excitation even in the presence of an underlying lattice which explicitly breaks translation invariance is a consequence of long range phase coherence in the SF. Its observation would thus constitute an independent proof that the atoms move coherently over the whole lattice and thus phase gradients give rise to dissipationless currents.

Finally we discuss the change in the atom number statistics at individual sites between the SF and the MI regime. This issue has been investigated in a beautiful experiment, observing collapse and revival of the matter wave due to the coherent superposition of states with different atom numbers in the SF [18]. As noted above, the ground state (12) in the extreme MI limit is a product of Fock states with a definite number  $\bar{n}$  of particles at each site. At finite hopping  $J \neq 0$ , this simple picture breaks down because the atoms have a finite amplitude to be at different sites. The many body ground state can then no longer be written as a simple product state as in (12). In the opposite limit  $U \rightarrow 0$ , the ground state is a condensate of zero quasimomentum Bloch states. In the limit  $N, M \rightarrow \infty$  at fixed (not necessarily integer) density  $N/M$ , the associated perfect condensate (13) becomes indistinguishable in a local measurement from a coherent state

$$|\Psi_{SF}\rangle(U=0) = \exp \sqrt{N} \hat{b}_{q=0}^\dagger |0\rangle = \prod_{l=1}^M \left( \exp \sqrt{\frac{N}{M}} \hat{b}_l^\dagger |0\rangle_l \right) \quad (17)$$

which factorizes into a product of *local* coherent states with average  $\langle \hat{n}_l \rangle = N/M$  because boson operators at different sites commute. The probability distribution for the number  $n$  of atoms at any given site in a perfect SF state in an optical lattice is therefore Poissonian. A remarkable consequence of the representation (17) is that, at least for integer densities  $N/M = \bar{n} = 1, 2, \dots$ , the many body ground state may be written in a local product form

$$|\Psi_{GW}\rangle = \prod_l \left( \sum_{n=0}^{\infty} c_n |n\rangle_l \right) \quad (18)$$

in both limits  $J \rightarrow 0$  and  $U \rightarrow 0$ . The associated atom number probability distribution  $p_n = |c_n|^2$  is either a pure Fock or a full Poissonian distribution. It is now very plausible to use the factorized form (18) as an *approximation* for arbitrary  $J/U$ , taking the coefficients  $c_n$  as variational parameters which

are determined by minimizing the ground state energy [19]. As pointed out by Rokhsar and Kotliar [20], this is effectively a Gutzwiller ansatz for bosons. Beyond being very simple computationally, this ansatz describes the SF to MI transition in a mean-field sense, becoming exact in infinite dimensions. In addition, it provides one with a very intuitive picture of the transition to a MI state, which occurs precisely at the point, where the local number distribution becomes a pure Fock distribution. This is consistent with a vanishing expectation value of the local matter wave field

$$\langle \Psi_{GW} | \hat{b}_l | \Psi_{GW} \rangle = \sum_{n=1}^{\infty} \sqrt{n} c_{n-1}^* c_n \quad (19)$$

in the Gutzwiller approximation. It is important, however, to emphasize that the ansatz (18) fails to account for the nontrivial correlations between different sites present at any finite  $J$ . These correlations imply that the one particle density matrix  $\rho_1(\mathbf{R})$  is different from zero at finite distance  $|\mathbf{R}| \neq 0$ , becoming long ranged at the transition to a SF. By contrast, in the Gutzwiller approximation, the one particle density matrix has no spatial dependence at all: it is zero at any  $|\mathbf{R}| \neq 0$  in the MI and is completely independent of  $\mathbf{R}$  in the SF. Moreover, in the Gutzwiller approximation, the phase transition is directly reflected in the local number fluctuations, with the variance of  $n_l$  vanishing throughout the MI phase. By contrast, in an exact theory local variables like the on-site number distribution will change in a smooth manner near the transition and the variance of the local particle number will only vanish in the limit  $J \rightarrow 0$ . Moreover, the Gutzwiller approximation fails to correctly describe long wavelength excitations [20]. Different methods are thus needed to discuss e.g. the dynamics of formation of a SF from an initial MI phase after a sudden reduction of the optical lattice depth [21,22].

### 3 Atomic Quantum Wires

For interacting fermions, one of the fundamental paradigms of many body theory is Landau's Fermi liquid concept. It states that the low lying excitations of any normal and fluid phase of fermions can be described by a set of fermionic quasiparticles, just as in an ideal Fermi gas. As is well known, Fermi liquid theory is invalid in 1D, where the generic interacting Fermi system is a Luttinger liquid [23]. As pointed out by Haldane [24], this concept is not confined to fermions but in fact applies to all interacting 1D quantum liquids, because statistics and interaction cannot be separated in one dimension. Now fermionic Luttinger liquids have been investigated intensively in the last years in condensed matter physics and spin-charge separation as one of their hallmark properties has been verified experimentally in semiconductor quantum wires [25]. By contrast, no realization of bosonic Luttinger liquids is known so far. In the following, we will show that cold atoms confined in a strong op-

tical lattice can be used to realize atomic quantum wires and allow to study Luttinger liquid physics in a bosonic system.

Consider a homogeneous Bose gas which is confined to an effectively 1D situation by a strong harmonic trapping potential. The associated excitation energy  $\hbar\omega$  is assumed to be much larger than the chemical potential and thus only the lowest transverse eigenmode is accessible. The relevance of interactions in the resulting single mode quantum wire may again be characterized by the ratio between the interaction and kinetic energy per particle [26]

$$\gamma = \frac{g_1 n}{\hbar^2 n^2 / m} = \frac{2a_s}{nl_\perp^2} \quad (20)$$

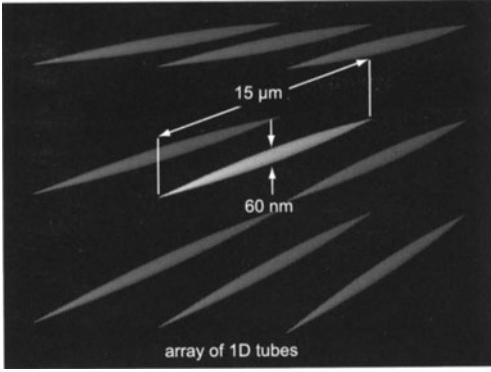
Here  $g_1$  is the strength of the effective delta-function interaction in 1D and  $l_\perp$  the oscillator length for the transverse confinement. Remarkably, the dimensionless interaction strength  $\gamma$  scales inversely with the 1D density  $n$ . In one dimension, therefore it is the *low* density limit where interactions dominate. This rather counterintuitive result can be understood physically by noting that at low 1D densities, the average kinetic energy per particle  $\epsilon_{kin} \sim n^2$  becomes negligible compared with the interaction energy. As a result, the atoms are perfectly reflected by the repulsive potential of the surrounding particles. For  $\gamma \gg 1$ , therefore, the system approaches a gas of impenetrable bosons which is called the Tonks-limit [26]. In particular at  $\gamma = \infty$  the exact many body wave function of  $N$  hard core bosons

$$\Psi_B(x_1 \dots x_N) = \prod_{i < j} |\sin [\pi(x_j - x_i)/L]| = |\Psi_F^{(0)}(x_1 \dots x_N)| \quad (21)$$

is just the absolute value of that of a free Fermi gas, as was shown a long time ago by Girardeau [27]. In order to describe the situation at arbitrary values of  $\gamma$ , it is convenient to use Haldane's description of 1D Bose liquids with short range interactions in terms of their long wavelength density oscillations [24]. Introducing canonically conjugate fields  $\hat{\phi}(x)$  and  $\hat{\theta}(x)$  such that  $[\partial_x \hat{\theta}(x), \hat{\phi}(y)] = i\pi\delta(x-y)$  and  $\hat{\theta}(x)$  related to the fluctuations  $\delta\hat{n}(x)$  around the average density via  $\delta\hat{n}(x) = \partial_x \hat{\theta}(x)/\pi$ , the effective Hamiltonian describing the low lying excitations is that of 1D noninteracting gas of phonons

$$\hat{H} = \frac{\hbar v_s}{2\pi} \int dx \left[ K(\partial_x \hat{\phi})^2 + \frac{1}{K} (\partial_x \hat{\theta})^2 \right] \quad (22)$$

with sound velocity  $v_s$ . The low energy physics of a 1D Bose liquid is thus completely determined by the velocity  $v_s$  and the dimensionless ratio  $K = \pi\hbar n/mv_s$ . The precise relation between the microscopic coupling constant  $\gamma$  and  $K$  may be obtained from the exact solution of the 1D short range Bose gas by Lieb and Liniger [28]. It turns out that  $K$  decreases monotonically from  $K(\gamma \rightarrow 0) \rightarrow \pi/\sqrt{\gamma} \gg 1$  in the weak interaction, high density limit to  $K(\gamma \gg 1) \rightarrow 1$  in the strong interaction, low density Tonks limit. The



**Fig. 3.** The geometry and size of 1D atomic quantum wires created by an optical lattice. The spacing between the wires in the horizontal and vertical direction is  $\lambda/2 = 413 \text{ nm}$ . Taken from [29]

parameter  $K$  determines the asymptotic behavior of all correlation functions, e.g. the power in the characteristic decay

$$\rho_1(x, T) \sim \left( \frac{\hbar v_s}{k_B T} \sinh \frac{\pi |x| k_B T}{\hbar v_s} \right)^{-1/2K} \quad (23)$$

of the one particle density matrix [24].

An experimental realization of single channel quantum wires filled with bosonic atoms has been achieved very recently by Moritz et.al. [29]. Using an optical lattice with two orthogonal retro-reflected laser beams which produce the standing wave, the atoms are only confined in the radial  $x, y$ -direction, while their motion along the axial  $z$ -direction is bound only by a weak magnetic trap. Loading an initially spherical cloud of condensed atoms into this lattice thus leads to an array of 1D tubes of atoms which are tightly confined in the radial direction and are separated by the periodicity of the lattice  $a = \lambda/2$ , as shown schematically in Fig. 3.

In order to verify that these wires are indeed true 1D quantum liquids, the authors have measured the frequency  $\omega_B$  of the lowest compressional mode of the atoms by exciting the system with a periodic modulation of the intensity of the optical lattice. As was shown by Menotti and Stringari [30] using sum rule arguments, the frequency  $\omega_B$  in a 1D trapped gas is directly related with the mean square size  $\langle z^2 \rangle$  of the cloud by

$$\omega_B^2 = -\frac{2\langle z^2 \rangle}{d\langle z^2 \rangle/d\omega_z^2}. \quad (24)$$

Here  $\omega_z$  is the small (in practice  $\omega_z = 2\pi \cdot 84 \text{ Hz}$ ) trapping frequency of the axial confinement. It is identical with the frequency  $\omega_D$  of the dipole mode which, by Kohn's theorem, is completely independent of the interaction. The frequency of the compressional mode  $\omega_B$ , in turn, gives information both about the dimensionality and the interaction strength. In a 3D weakly

interacting BEC one has  $(\omega_B/\omega_z)^2 = 5/2$  [1] while in 1D a different ratio is obtained. To see this, we use the standard local density approximation

$$\mu[n(z=0)] = \mu_{hom}[n(z)] + \frac{1}{2}m\omega_z^2 z^2 \quad (25)$$

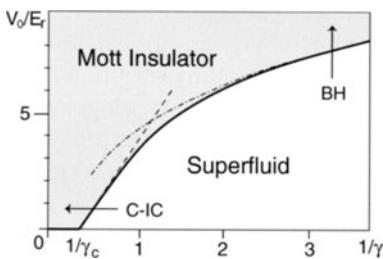
for the density profile  $n(z)$  in a trap. In the weakly interacting regime, the chemical potential of the homogeneous system is given by the mean field expression  $\mu_{hom}[n] = g_1 n$ . The resulting density profile is therefore an inverted parabola and the associated Thomas-Fermi radius  $R$  scales like  $R \sim \omega_z^{-2/3}$  at fixed total particle number  $N$ . With  $\langle z^2 \rangle \sim R^2$  equation (24) therefore immediately gives a ratio  $(\omega_B/\omega_z)^2 = 3$ , different from the 3D situation. A precise measurement of  $\omega_B$ , thus gives a clear signature for a truly 1D quantum liquid of bosons [29]. At the lowest attainable densities, the interaction parameter  $\gamma$  reaches values of order one, i.e. one is essentially in the crossover regime to the strongly interacting limit. As noted above, for  $\gamma \gg 1$ , the 1D Bose liquid is expected to behave like a fermionic system, where  $\mu[n] \sim n^2$ . In this Tonks regime, the characteristic radius  $R$  of the cloud at fixed  $N$  therefore scales like  $R \sim \omega_z^{-1/2}$ , leading to a frequency ratio  $(\omega_B/\omega_z)^2 = 4$ . In general, the density profile changes continuously between the mean field and the Tonks regime [32]. A measurement of the frequency ratio  $\omega_B/\omega_z$  thus should allow to observe the continuous change from a Bose to a Fermi gas in a 1D interacting system of bosons as predicted by Lieb and Liniger [28] and Haldane [24]. While the Tonks regime has not yet been realized experimentally, the most recent progress in this area [31] has investigated the crossover from 3D to 1D behavior in the SF-MI transition. This is achieved using an optical lattice with different amplitudes  $V_\perp$  and  $V_\parallel$  of the periodic potential felt by the atoms in the radial  $x, y$ - and axial  $z$ -direction. Choosing  $V_\perp = 30 E_r$ , hopping in the transverse direction is negligible and the system behaves like an array of 1D atomic wires. By continuously lowering  $V_\perp$ , it is then possible to observe the crossover from a 1D to a 3D situation. In particular the corresponding change in the Mott-Hubbard transition can be monitored by measuring the excitation spectrum via periodically modulating the amplitude of the axial lattice potential. This allows a direct spectroscopic observation of both the continuum of excitations in the SF regime and the excitation gap in the Mott phase [31].

Clearly experiments of this kind open a unique possibility of studying strong correlations in any dimension. As one example, we want to discuss what happens in a 1D Bose gas in a *weak* optical lattice, where a description in terms of the Bose-Hubbard model is no longer possible. As pointed out above, in 1D the dimensionless interaction strength  $\gamma$  increases with decreasing density. In the strong coupling regime the superfluid ground state of the homogeneous system is therefore expected to be destroyed already in a weak optical lattice. In fact, it turns out that this problem can be solved essentially analytically [33]. The basic idea is that a strongly interacting 1D Bose liquid

behaves like a noninteracting Fermi gas, as shown by Girardeau [27]. Moreover, this equivalence remains valid in the presence of an arbitrary additional one-particle potential like that of an optical lattice. For a qualitative understanding of a strongly interacting 1D Bose liquid in a weak optical lattice, it is therefore useful to consider a free Fermi gas in a weak periodic potential  $V(x) = V_0 \sin^2 kx$  with  $V_0$  of order  $E_r$  or smaller. This is an elementary problem in solid state physics, equivalent to the nearly free electron limit of a 1D bandstructure. The single particle spectrum consists of a series of free particle like bands separated by energy gaps  $\Delta_l$   $l = 1, 2, \dots$ . The gaps become exponentially small with increasing energy, scaling like  $\Delta_l \sim |V_0|^l$  in the limit  $V_0 \ll E_r$ . For a commensurate density, where an integer number  $l$  of particles fit into one unit cell, the  $l$  lowest bands are completely filled. The groundstate of noninteracting fermions is thus a trivial band insulator. Similar to the incompressible Mott-insulating phase of the Bose-Hubbard model, the state with a fixed integer density remains locked over a finite range  $\Delta_l$  of the chemical potential. For weak optical lattices  $V_0 \ll E_r$  the lowest gap  $\Delta_{l=1} = |V_0|/2$  is much larger than the higher order ones. As a result, it is the commensurate phase with unit filling  $N = M$  which has maximal stability. In the Tonks limit  $\gamma \rightarrow \infty$  an arbitrary weak optical lattice commensurate with the average density will therefore pin the atoms into an incompressible MI phase. In order to discuss whether this peculiar feature of hard core bosons in one dimension is still present at finite and experimentally accessible values of  $\gamma$ , it is convenient to use Haldane's description (22) of 1D bosons in terms of their long wavelength density oscillations. Adding a periodic potential commensurate with the average particle density gives rise to an additional nonlinear term  $\cos 2\theta(x)$ , leading to the well known sine-Gordon model [23]. This model is exactly soluble and exhibits a transition at a critical value  $K_c = 2$  such that for  $K > 2$  the ground state remains gapless and superfluid in a weak optical lattice while for  $(1 <)K < 2$  the atoms are locked even in an arbitrary weak periodic lattice. The commensurate phase is characterized by a finite excitation gap [33]

$$\Delta = \frac{2E_r}{K} \left( \frac{K|V_0|}{(2-K)4E_r} \right)^{1/(2-K)} \quad (26)$$

which is nonanalytic in  $V_0$ . It approaches  $|V_0|/2$  in the Tonks gas limit  $K \rightarrow 1$  in agreement with the ideal Fermi gas picture discussed above. Using the exact result for  $K(\gamma)$  which follows from the Lieb and Liniger solution of the 1D Bose gas, it turns out that the critical value  $K_c = 2$  is reached at  $\gamma_c \approx 3.5$ . The transition to a commensurate, incompressible state in a weak optical lattice thus occurs in a regime which is not far from that, already achieved in the atomic wires discussed above. Of course, as equation (26) shows, the gap is exponentially small in the vicinity of the critical value  $K_c = 2$ . An experimental observation therefore requires optical lattice depths  $V_0$  which are of the order of the recoil energy  $E_r$ . For larger values of  $V_0$  one eventually



**Fig. 4.** Ground state phase diagram of a one-dimensional Bose gas in an optical lattice of arbitrary strength  $V_0/E_r$  as a function of the inverse interaction parameter  $\gamma$ . The dashed line is the asymptotic behaviour near the critical point  $\gamma_c$  as determined from the sine-Gordon model, while the dashed-dotted line derives from the solution of the transcendental equation (27) obtained from the critical point of the 1D Bose-Hubbard model

has to recover the physics of the Bose-Hubbard model. This limit is reached in the weak coupling regime  $\gamma \ll 1$ , where the SF-MI transition requires a deep optical lattice. Calculating the band width parameter  $J$  and the effective on site repulsion  $U$  in a 1D situation in terms of the dimensionless optical lattice depth  $V_0/E_r$ , the criterion  $U/J|_c = 2C \approx 3.37$  for the transition in the 1D Bose-Hubbard model can be written in a form similar to equation (14)

$$\frac{4V_0}{E_r} = \ln^2 \left[ 4\sqrt{2}\pi C(V_0/E_r)^{1/2}/\gamma \right]. \quad (27)$$

The solution of this transcendental equation gives a critical value  $V_0/E_r|_c$  which increases rather slowly as a function of the inverse interaction parameter  $1/\gamma$ . The resulting phase boundary matches smoothly to the one obtained for a weak optical lattice from the known Kosterlitz-Thouless nature of the transition near  $K_c = 2$ , as shown in Fig. 4. The Bose-Hubbard transition (BH) for weakly interacting gases in a deep optical lattice thus continuously evolves into one of the commensurate-incommensurate (C-IC) type for strongly interacting gases in a weak lattice. An extension of these results to a nonvanishing transverse hopping in an array of 1D quantum wires has very recently been given by Ho, Cazalilla and Giamarchi [34].

## 4 Perspectives

As the examples above clearly show, cold atoms provide a novel tool to study the physics of strong correlations in both a widely tunable range and in unprecedentedly clean systems. Basic models in many body theory like the Hubbard-model for bosons or fermions with on-site interaction, which were originally introduced in a condensed matter context as an idealized description of strong correlation effects in real materials can now be applied on a quantitative level. Beyond Mott-Hubbard and Luttinger liquid physics discussed above, there are a number of other strong correlation phenomena which appear to be realizable with bosonic atoms. These include spin-dependent optical lattice potentials which depend on the internal state

of the atoms. They may be used e.g. to change the sign of the tunneling amplitude  $J$  from positive to negative values [35] or to realize an analog of a quantum dot with a tunable coupling to the environment [36], as an example of a quantum impurity problem with cold atoms. Loading optical lattices with bosons in different internal states ('spinor condensates') might allow to generate exotic phases with fractionally charged excitations [37] or study fermionic Hubbard physics in a 1D two-component Bose system [38]. Also the influence of interactions on Anderson localization in bosonic systems and the related Bose-glass phase [13,39] may in principle be studied by placing a speckle pattern on top of an optical lattice [40]. Of course, beyond bosons, future interest will also focus on degenerate fermions. The recent realization of a BCS-BEC crossover with a Feshbach resonance [6,7,8] including clear indications for pairing on the BCS-side of the crossover [41], opens the way to study this basic phenomenon also in optical lattices or in 1D, where an exact solution of the crossover is possible [42]. With repulsive fermions in 1D, the equivalent of spin-charge separation appears accessible [43]. A further quite promising development are Bose-Fermi mixtures in optical lattices, in which supersolid phases, counterflow superfluidity or composite fermions are predicted to occur [44,45,46]. Finally, from a condensed matter point of view, one of the most challenging problems is of course the realization and study of the fermionic repulsive Hubbard model in an optical lattice, in particular in 2D. As discussed by Hofstetter et.al. [47] such a system would constitute a cold atom version of one of the most intensely studied models in condensed matter physics, allowing to access antiferromagnetic and unconventional normal and superconducting phases as found in the high temperature superconductors. Of course, realization of these models with cold atoms would not solve the latter problem, however it would be an extremely valuable tool to test some of the still open issues in this field. It is obvious that many of these proposals are quite challenging and their realization may take some time. However in view of the tremendous progress in this field in the last years, the perspectives look very encouraging and it is obvious that cold atoms will be a major field of research in many body physics in the future.

## References

1. F.D. Dalfovo, S. Giorgini, L.P. Pitaevski, S. Stringari, Rev. Mod. Phys. **71**, 463 (1999).
2. A.J. Leggett, Rev. Mod. Phys. **73**, 307 (2001).
3. B. DeMarco, D.S. Jin, Science **285**, 1703 (1999).
4. Y. Castin, R. Dum, Phys. Rev. A **57**, 3008 (1998).
5. E. Timmermanns, P. Tommasini, M. Hussein, A. Kerman, Physics Reports **315**, 199 (1999).
6. M. Greiner, C.A. Regal, D.S. Jin, Nature **426**, 537 (2003).
7. S. Jochim, M. Bartenstein, A. Altmeyer, G. Hendl, S. Riedl, C. Chin, J. Hecker Denschlag, R. Grimm, Science **302**, 2101 (2003).

8. M.W. Zwierlein, C.A. Stan, C.H. Schunk, S.M.F. Raupach, S.Gupta, Z. Hadzibabic, W. Ketterle, Phys. Rev. Lett. **91**, 250401 (2003).
9. P.O. Fedichev, M.W. Reynolds, G. V. Shlyapnikov, Phys. Rev. Lett. **77**, 2921 (1996).
10. E. A. Donley, Nature **417**, 529 (2002).
11. M. Greiner, O. Mandel, T. Esslinger, T. Hänsch, I. Bloch, Nature **415**, 39 (2002).
12. D. Jaksch, C. Bruder, J.I. Cirac, C.W. Gardiner, P. Zoller, Phys. Rev. Lett. **81**, 3108 (1998).
13. M.P.A. Fisher, P.B. Weichmann, G. Grinstein, D.S. Fisher, Phys. Rev. **B 40**, 546 (1989).
14. T.D. Kühner, S. White, H. Monien, Phys. Rev. B **61** 12474 (2000)
15. V.A. Kashurnikov, N.V. Prokof'ev, B.V. Svistunov, Phys. Rev. A **66**, 031601(R) (2002)
16. R. Roth R, K. Burnett, Phys. Rev. A **67**, 031602 (R) (2003)
17. C. Kollath, U. Schollwöck, J. von Delft, W. Zwerger, Phys. Rev. A **69**, 031601 (2004).
18. M. Greiner, O. Mandel, T.W. Hänsch, I. Bloch, Nature **419**, 51 (2002).
19. K. Sheshadri, H.R. Krishnamurty, R. Pandit, T.V. Ramakrishnan, Europhys. Lett. **22**, 257 (1993).
20. D.S. Rokhsar, B.G. Kotliar, Phys. Rev. B **44**, 10328 (1991).
21. E. Altman, A. Auerbach, Phys. Rev. Lett. **89**, 250404 (2002).
22. A. Polkovnikov, S. Sachdev, S.M. Girvin, Phys. Rev. A **66**, 033609 (2003).
23. H.J. Schulz, *Fermi liquids and non-Fermi liquids* in: *Proceedings of Les Houches Summer School LXI*, (Elsevier 1995); cond-mat/9503150.
24. F.D.M. Haldane, Phys. Rev. Lett. **47**, 1840 (1981).
25. Y. Tserkovnyak, B.I. Halperin, O.M. Auslaender, A. Yacoby, cond-mat/0312159.
26. D.S. Petrov, G.V. Shlyapnikov, J.T.M. Walraven, Phys. Rev. Lett. **85**, 3745 (2000).
27. M. Girardeau, J. Math. Phys. **1**, 516 (1960).
28. E.H. Lieb, W. Liniger, Phys. Rev. **130**, 1605 and 1616 (1963).
29. H. Moritz, T. Stöferle, M. Köhl, T. Esslinger, Phys. Rev. Lett. **91**, 250402 (2003).
30. C. Menotti, S. Stringari, Phys. Rev. A **66**, 043610 (2002).
31. T. Stöferle, H. Moritz, C. Schori, M. Köhl, T. Esslinger, Phys. Rev. Lett. **92**, 130403 (2004).
32. V. Dunjko, V. Lorent, M. Olshanii, Phys. Rev. Lett. **86**, 5413 (2001).
33. H.P. Büchler, G. Blatter, W. Zwerger, Phys. Rev. Lett. **90**, (2003).
34. A.F. Ho, M.A. Cazalilla, T. Giamarchi, cond-mat/0310382.
35. K.V. Krutitsky, R. Graham, Phys. Rev. Lett. **91**, 240406 (2003).
36. A. Recati, P.O. Fedichev, W. Zwerger, J. von Delft, P. Zoller, wnd-mat/0404533.
37. E. Demler, F. Zhou, Phys. Rev. Lett. **88**, 163001 (2002).
38. B. Paredes, J.I. Cirac, Phys. Rev. Lett. **90**, 150402 (2003).
39. S. Rapsch, U. Schollwöck, W. Zwerger, Europhys. Lett. **46**, 559 (1999).
40. B. Damski, J. Zakrzewski, L. Santos, P. Zoller, M. Lewenstein, Phys. Rev. Lett. **91**, 080403 (2003).
41. C.A. Regal, M. Greiner, D.S. Jin, Phys. Rev. Lett. **92**, 040403 (2004).

42. J.N. Fuchs, A. Recati, W. Zwerger, cond-mat/0402448.
43. A. Recati, P.O. Fedichev, W. Zwerger, P. Zoller, Phys. Rev. Lett. **90**, 020401 (2003).
44. H.P. Büchler, G. Blatter, Phys. Rev. Lett. **91**, 130404 (2003).
45. A.B. Kuklov, B.V. Svistunov, Phys. Rev. Lett. **90**, 100401 (2003).
46. M. Lewenstein, L. Santos, M.A. Baranov, H. Fehrmann, Phys. Rev. Lett. **92**, 050401 (2004).
47. W. Hofstetter, J.I. Cirac, P. Zoller, E. Demler, M.D. Lukin, Phys. Rev. Lett. **89**, 220407 (2002).

## Part VI

### **Thin Films and Surfaces**

# Plasma-Assisted Deposition and Crystal Growth of Thin Indium-Tin-Oxide (ITO) Films

R. Hippler<sup>1</sup>, H. Steffen<sup>1</sup>, M. Quaas<sup>2</sup>, T. Röwf<sup>1</sup>, T.M. Tun<sup>1</sup>, and H. Wulff<sup>2</sup>

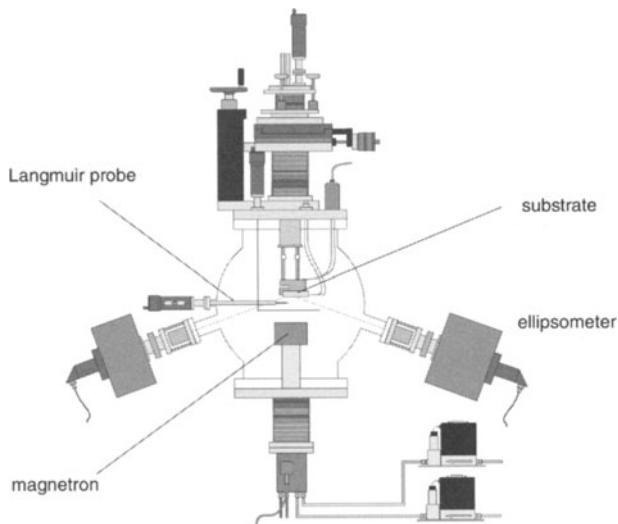
<sup>1</sup> Institut für Physik, Ernst-Moritz-Arndt-Universität Greifswald  
Domstraße 10a, 17489 Greifswald, Germany  
[Hippler@Physik.Uni-Greifswald.de](mailto:Hippler@Physik.Uni-Greifswald.de)

<sup>2</sup> Institut für Chemie und Biochemie, Ernst-Moritz-Arndt-Universität Greifswald  
Soldmannstr. 16, 17489 Greifswald

**Abstract.** Metal oxide films are of significant technological relevance and play an important role, e.g., as protective coatings and in the fabrication of solar cells. Plasma-assisted deposition employing magnetron discharges are frequently employed. Formation and crystal growth of, e.g., indium tin oxide (ITO) films show pronounced dependencies on plasma parameters and on the applied substrate voltage. The influence on chemical composition, crystal structure, density, etc., as well as possible reasons will be discussed.

## 1 Introduction

Metal oxide films are of significant technological relevance and play an important role, e.g., as protective coatings, in the fabrication of solar cells, and in plasma display panels. Thin films of, e.g., tin-doped indium oxide (indium tin oxide, ITO) are characterized by low electrical resistance and high optical transmittance in the visible region. The combination of these properties makes ITO films indispensable in many optoelectronic devices. The macroscopic properties of thin films are directly associated with their chemical composition, microstructure, and defect structure. In plasma-assisted deposition, film properties are to a large extent determined by plasma parameters. Therefore the correlation between deposition conditions and film composition and microstructure is of fundamental interest. Reactive direct current (DC) magnetron sputtering has been reported to yield ITO films with low resistance and good optical transmittance [1]. However, the influence of the deposition conditions on microscopic or structural properties of reactive sputtered ITO films has not been studied in detail, yet. Two deposition procedures are commonly in use; the reactive sputtering from a metallic In/Sn target in the presence of oxygen or from an indium-tin-oxide target. In these complex processes the energetic and thermal conditions at the substrate surface and the growing film surface play a dominant role. Elementary processes like adsorption, desorption, and diffusion as well as chemical reactions will be affected. Stoichiometry, microstructure and morphology of the deposited



**Fig. 1.** Deposition chamber with magnetron (schematic)

films depend on the energetic conditions at the surface. Only recently, more systematic investigations on the role of plasma parameters ion flux and ion energy to the substrate have been reported [2,3]. Many authors use oxygen as reactive gas during the deposition procedure from oxide targets to form stoichiometric ITO; it is well known, that the oxygen partial pressure is a decisive parameter during ITO film deposition by magnetron sputtering [4]. For example, Mergel et al. [5] observed the formation of hyper-stoichiometric ITO films deposited at high oxygen flows. Also, applying a negative substrate voltage during film deposition is rather common. Thus, energy and flow density of charged particles (electrons, ions) to the substrate can be varied. In recent investigations we found that negative substrate voltage influences the composition of the growing film [6,7]. The present study focuses on the relations between the deposition conditions at different oxygen flows and negative substrate voltages and the influence on the chemical composition and the microstructure of ITO films. The plasma in front of the substrate was investigated by energy resolved mass spectrometry. The energy influx to the substrate was estimated by a thermoprobe. Grazing incidence x-ray diffractometry (GIXD) and photoelectron spectroscopy (XPS) were applied to determine the chemical composition and the phase composition of the deposited films. Film thickness and density were measured by grazing incidence x-ray reflectometry (GIXR).

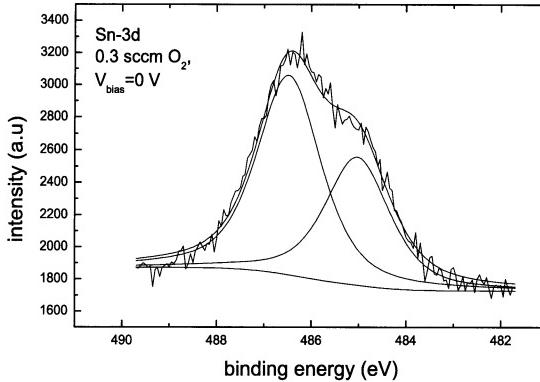
## 2 Experiment

Thin ITO films were deposited on Si(100) wafers by DC planar magnetron sputtering (Figure 1). A metallic In/Sn (90/10) target was used. Residual gas pressure was  $< 10^{-8}$  mbar. Deposition pressure ( $5.6 \times 10^{-3}$  mbar), gas flow (argon 15 sccm) and discharge power (30 W) were kept constant. The flow of the reactive oxygen gas was varied between 0 and 2 sccm by means of a mass flow controller (MKS Instruments). Substrate voltages  $U_{sub}$  between 0 and  $-100$  V were applied. ITO layers were deposited from a metallic and from an oxidized target. To obtain a metallic target, the target was cleaned by sputtering for about 30 minutes with pure argon as working gas. To get an oxidized target we worked for at least 30 minutes with additional 2 sccm oxygen gas flow prior to deposition.

X-ray photoelectron spectroscopy (XPS) measurements were performed with a conventional surface analysis system (VG MT 500) to investigate the composition of the deposited ITO films. Samples were transferred under vacuum from the deposition chamber to the surface analysis chamber. Figure 2 shows a typical XPS result for the Sn  $3d_{5/2}$  peak obtained using the Mg K $\alpha$  x-ray line ( $h\nu = 1253.6$  eV), and its decomposition into two, metallic tin and tin oxide, peaks.

Grazing incidence x-ray reflectometry (GIXR) investigations were performed on a  $\theta - 2\theta$  diffractometer (Siemens D 5000) with a special thin film attachment [8]. Film thickness and density were determined by fitting the resulting reflectometry curves using the REFSIM1.1 software (Siemens). The deposition time for samples investigated by GIXR was 30 s. Knowing the deposition rate for every deposition condition, a second series of samples was produced with films of 50 nm thickness. These films were investigated by grazing incidence x-ray diffractometry (GIXD) to investigate the phase composition in the films. The angle of the incident x-ray beam was fixed at  $0.7^\circ/2\theta$ . The scanning region was  $25-45^\circ/2\theta$  (step width  $0.02^\circ/2\theta$ , scanning time 2 s per step). The as-deposited samples consist of crystalline metallic In/Sn and/or x-ray amorphous indium-tin-oxide depending on the deposition conditions.

Crystalline indium-tin-oxide films were obtained by isothermal post-annealing at  $200$  °C in a high temperature chamber (Bühler HDK2.4) attached to a  $\theta - \theta$ -diffractometer (Seifert XRD 3000). During annealing the films were step-scanned continuously at the measuring conditions mentioned above. For every measurement the integral intensity of the ITO(222) x-ray reflection was determined by a pseudo-Voigt fit procedure. From the time dependence of the integral ITO(222) x-ray intensity during annealing the diffusion coefficients  $D$  and the initial amount of amorphous indium-tin-oxide in the as-deposited films were determined. The amount of amorphous ITO in the as-deposited films is characterized by the initial phase composition factor  $f_0$  which we define as the concentration ratio of indium oxide relative to the



**Fig. 2.** Sn-3d<sub>5/2</sub> XPS peak of deposited indium-tin-oxide and its decomposition (see text)

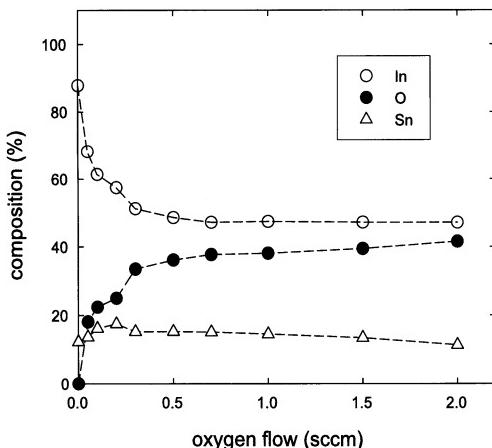
total indium concentration in the film. The mathematical model used here is described in detail in Ref. [9].

The energy influx from the plasma towards the substrate has been measured by a simple thermoprobe [10,11]. The probe consists of a copper substrate (diameter: 3.4 cm, thickness 0.02 cm), which is spot-welded to a thermocouple (type j) and placed within a solid shield. The substrate is only connected to the thermocouple and an additional wire for biasing. The measurement of the total energy influx is based on the determination of the difference between the time derivatives of the substrate temperature during heating (plasma on) and cooling (plasma off). This difference is proportional to the energy influx. The proportionality factor is the specific heat of the substrate. It was determined by a known thermal power source.

The energy distribution function IEDF of plasma ions was investigated by energy resolved mass spectrometry employing a so-called plasma monitor (Balzers PPM 421). Mainly positive Ar<sup>+</sup>, In<sup>+</sup>, atomic O<sup>+</sup> and molecular O<sub>2</sub><sup>+</sup> ions and negative O<sup>-</sup> ions were detected.

### 3 Results

In the following we report results for thin ITO film deposition as function of oxygen gas flow and substrate bias voltage. The deposited films were analyzed by x-ray photoelectron spectroscopy (XPS) and grazing incidence diffractometry (GIXD) and reflectometry (GIXR).



**Fig. 3.** The composition (○, In; △, Sn; ●, O) of deposited indium-tin-oxide films as function of oxygen flow for a metallic target

### 3.1 Chemical Composition Determined by Photoelectron Spectroscopy (XPS)

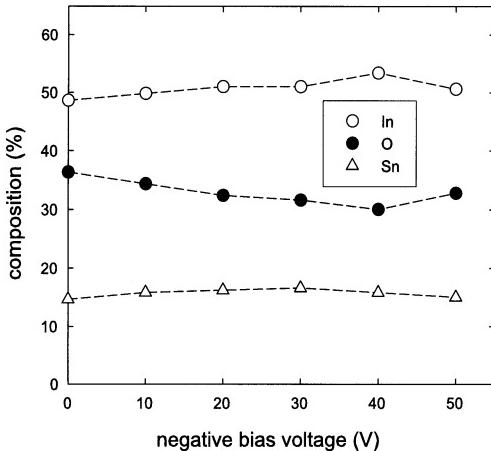
#### 3.1.1 Oxygen Dependence

Figure 3 shows the composition of thin ITO films deposited in Si(100) as function of oxygen gas flow, with the film composition investigated by XPS. Without oxygen flow, metallic indium/tin is deposited. With increasing oxygen flow, the amount of oxygen in the deposited films increases, while the indium content decreases. Above an oxygen flow of 0.5 sccm a saturation of the oxygen and indium contents occurs. The tin content varies between 11 % and 17 % and is, hence, significantly enhanced compared to the stoichiometric composition of the employed In/Sn target. Contrary to indium, the tin content also shows a pronounced increase with increasing oxygen flow, reaching a maximum around 0.2 sccm followed by a gradual decrease.

The above mentioned behavior indicates formation of an oxygen-poor indium-tin-oxide film, well below the stoichiometric oxygen fraction of 0.6 for  $\text{In}_2\text{O}_3$ . Further and more detailed studies reveal the existence of 2 phases, a metallic (oxygen-free) and a completely oxidized ITO phase.

#### 3.1.2 Bias Voltage Dependence

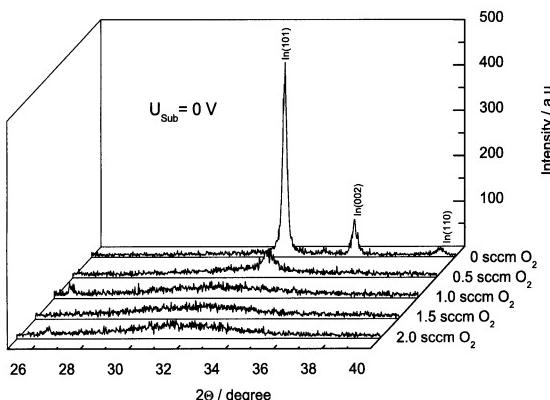
In addition, the composition of the deposited ITO films shows some variation with substrate bias voltage. Fig. 4 shows that the In content increases for negative bias voltages, reaching a maximum composition of about 54 % around  $-40$  V. Similarly, the Sn content increases from 14 → 16 %, whereas the oxygen content decreases from about 36 % → 30 %. Apparently, a negative bias voltage leads to a more metal-rich and oxygen-poor indium-tin-oxide film.



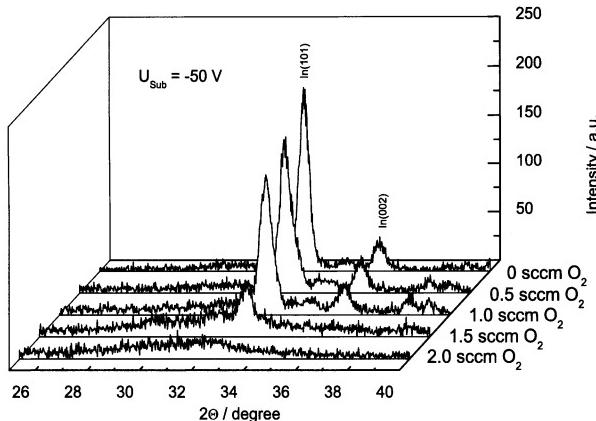
**Fig. 4.** The composition (○, In; △, Sn; ●, O) of deposited indium-tin-oxide films as function of negative bias voltage for a metallic target. Oxygen flow was 0.5 sccm

### 3.2 Film Properties Determined by GIXD

Figures 5 and 6 show the x-ray patterns of as-deposited layers. The films were deposited with different oxygen flows and with a metallic target. The layers shown in Fig. 5 were deposited with a substrate voltage  $U_{sub} = 0$  V, the layers of Fig. 6 with  $U_{sub} = -50$  V. Samples deposited at low oxygen flow show reflections attributable to polycrystalline metallic indium [12]. No crystalline indium oxide phases were detected. Crystalline Sn or  $\text{SnO}_x$  phases were also not observed. The latter is not surprising since Sn atoms are believed to substitute In atoms without changing the lattice structure [8]. The amount of crystalline metallic In/Sn decreases with increasing oxygen flow. In case of a grounded substrate ( $U_{sub} = 0$  V) the films apparently become x-ray amorphous for oxygen flows larger than 1 sccm. Applying a negative substrate voltage  $U_{sub} = -50$  V results in partly crystalline metallic films at larger oxygen flows. It appears, hence, that a negative bias voltage has the same



**Fig. 5.** GIXD patterns of as-deposited films for various oxygen flows ( $U_{sub} = 0$  V)

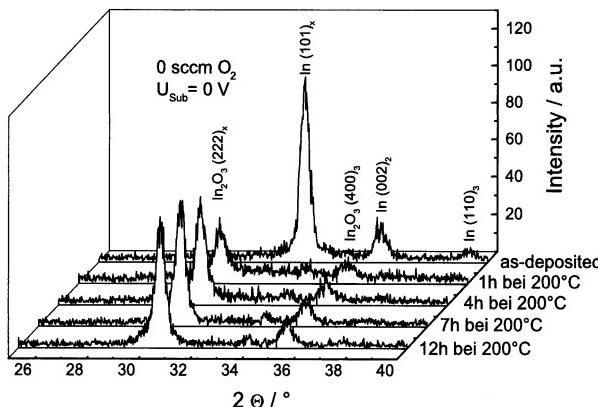


**Fig. 6.** GIXD patterns of as-deposited films for various oxygen flows ( $U_{sub} = -50\text{ V}$ )

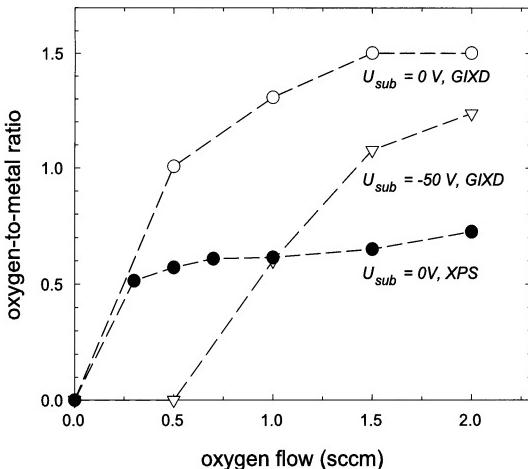
effect as reducing the oxygen flow to the substrate. The here observed effect is similar although much more pronounced compared to the bias voltage dependence noted already by XPS investigations (see above).

### 3.3 Post-annealing of Deposited Indium-Tin-Oxide Films

In order to extract further information the as-deposited films have been post-annealed for several hours at  $200^\circ\text{C}$ . During annealing the crystal structure changes from an amorphous to a crystalline indium-tin-oxide phase (Fig. 7). The appearance of a crystalline indium-tin-oxide phase is due to a rapid crystallization process of the initially amorphous indium-tin-oxide, and, if not enough oxygen is present in the film, by relatively slow diffusion of oxygen into



**Fig. 7.** GIXD patterns of post-annealed films for various annealing conditions. Film deposition was without oxygen flow and without substrate voltage



**Fig. 8.** Oxygen-to-metal ratio *versus* oxygen flow. Data are obtained from GIXD ( $\circ$ ,  $U_{sub} = 0\text{ V}$ ;  $\triangle$ ,  $U_{sub} = -50\text{ V}$ ), and from XPS ( $\bullet$ ,  $U_{sub} = 0\text{ V}$ ) (see text)

the film followed by crystallization. A model that separates the two oxygen contributions by their different time dependencies has been applied [9]; it allows to calculate the initial oxygen-to-metal ratio and the oxygen diffusion coefficient  $D$ . The model ignores the chemical difference between Sn and In atoms and, hence, treats Sn and In equally.

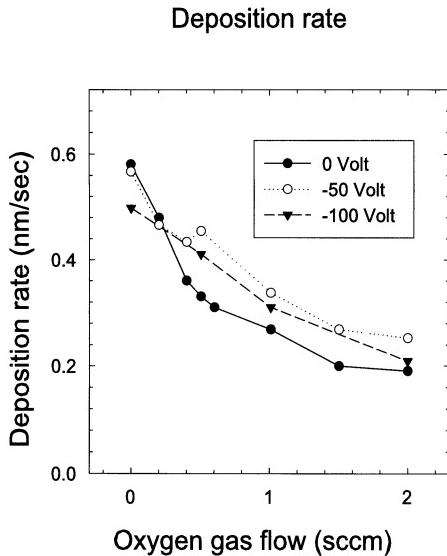
As is expected, the extracted oxygen-to-metal ratio shows a pronounced dependence on oxygen flow. There exists a distinct difference between the GIXD and the XPS measurements (Fig. 8). It seems to imply that deposited films consist of two phases: an amorphous metallic In phase not detected by GIXD and a fully oxidized indium-tin-oxide phase.

### 3.4 Deposition Rate and Film Density Determined by GIXR

The measured deposition rates and film densities derived from GIXR investigations are displayed in Figs. 9 and 10. With increasing oxygen flow the deposition rate (in nm/s) decreases by almost a factor of 3. At the same time, the film density increases, reaching the  $\text{In}_2\text{O}_3$  density for oxygen flows larger than about 1 sccm.

### 3.5 Oxidized Target

Indium-tin-oxide films deposited with an oxidized target differ significantly from those obtained with a metallic target. Fig. 11 shows the XPS results for as-deposited films (i.e., films that have been kept under vacuum), films that were exposed to ambient air, and post-annealed films. The as-deposited films now show the proper stoichiometric ratios of ITO, i.e., an oxygen content of about 60 %. The tin content is about 5 % and thus in good agreement with

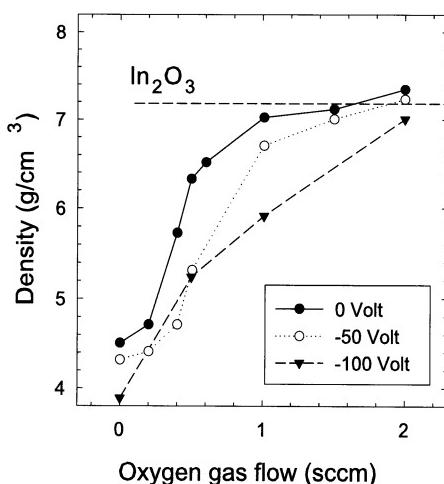


**Fig. 9.** Deposition rate of indium-tin-oxide films *vs.* oxygen flow for a metallic target and for the indicated substrate bias voltages

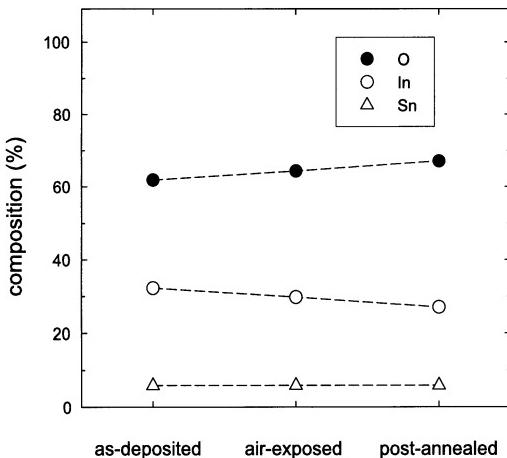
the stoichiometric ratio of the target. Air-exposed and post-annealed films show even higher oxygen contents, indicating oxygen-rich indium-tin-oxide. Similar observations have been reported by Mergel et al. [5] and explained by an incorporation of additional oxygen in the bixbyite structure, possibly into constitutional vacancies.

### 3.6 Ion Energy Distribution Function (IEDF)

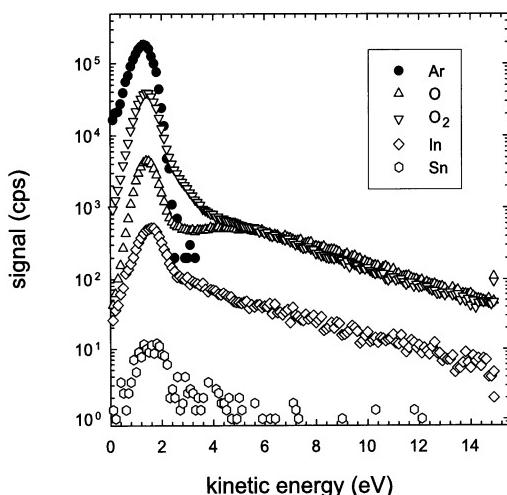
The observed dependencies of film characteristics on substrate voltage suggest that impinging ions play a significant role during film deposition. In order to



**Fig. 10.** Density of deposited indium-tin-oxide films *vs.* oxygen gas flow for a metallic target and for the indicated bias voltages



**Fig. 11.** Composition of as-deposited, air-exposed, and post-annealed indium-tin-oxide films ( $\circ$ , In; ●, O;  $\triangle$ , Sn)



**Fig. 12.** Kinetic energy distribution of positive ions in a dc magnetron discharge (argon gas flow 20 sccm, oxygen gas flow 1 sccm). ●, Ar<sup>+</sup>;  $\triangle$ , O<sup>+</sup>;  $\nabla$ , O<sub>2</sub><sup>+</sup>;  $\diamond$ , In<sup>+</sup>; ○, Sn<sup>+</sup>

investigate this influence we measured the kinetic energy distribution functions of positive and negative ions. Figure 12 shows the energy distribution function IEDF of positive ions measured by energy resolved mass spectrometry. Without oxygen gas flow the positive ion mass spectrum mainly consists of Ar<sup>+</sup> ions. The Ar<sup>+</sup> signal displays a maximum around 1 eV followed by a pronounced intensity drop. Note that the Ar<sup>+</sup> energy distribution displayed in Figure 12 was derived from the measured Ar-36 isotope signal and divided by the isotope abundance (0.3365 %) to yield the total Ar<sup>+</sup> signal.

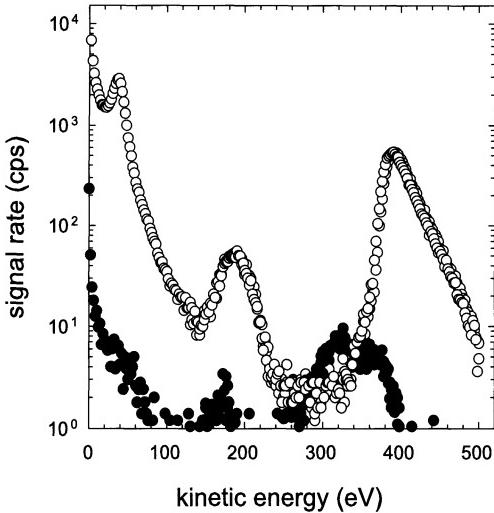
The energy distribution of  $\text{In}^+$  ions differs from this behavior as it displays a pronounced tail towards larger energies, presumably reflecting the physical sputtering process by which indium atoms are produced. The measured IEDF also shows that  $\text{In}^+$  ions only partly thermalize at the gas densities of interest here. The total  $\text{In}^+$  intensity without oxygen flow is about 2 orders of magnitude smaller than compared to  $\text{Ar}^+$ . With an oxygen flow of 1 sccm oxygen flow, the  $\text{In}^+$  signal reduces by a factor of 4, while only a modest reduction is observed for  $\text{Ar}^+$  ions. The oxygen ion intensity is now reaching a level comparable to the  $\text{Ar}^+$  signal. Most abundant are molecular  $\text{O}_2^+$  ions while atomic  $\text{O}^+$  ions are about one order of magnitude less abundant. Oxygen ions also display a high-energy tail suggesting that a certain fraction of these ions stems from sputtering processes at the magnetron target. Due to the negative biasing of typically  $-400\text{ V}$  it is unlikely, however, that positive ions directly escape from the target.

Opposite to positive ions which become accelerated in the plasma sheath region, negative ions are partly trapped inside the plasma by the positive plasma potential of typically 1-2 Volt. Only negative ions with sufficient kinetic energy can overcome this potential barrier. The most abundant negative ion species is  $\text{O}^-$ . The energy distribution function of negative  $\text{O}^-$  ions consists of several groups. Most intense is a low energy group with near-zero kinetic energies. These ions are believed to be produced inside the plasma. Several groups of high-energy  $\text{O}^-$  exist. Most prominent is a group of ions around  $50\text{ eV}$  followed by another small group around  $200\text{ eV}$ . The last group has energies of  $400\text{--}500\text{ eV}$ ; its origin is due to the sputtering of oxygen atoms at the target. This is corroborated by its sharp onset which corresponds to the voltage drop at the target. The origin of the first and second groups of the high-energy ions is presumably also due to sputtering of atomic or molecular oxygen atoms at the target. Negative molecular  $\text{O}_2^-$  ions were also observed, albeit at a significantly reduced total intensity of few percent compared to the total intensity of atomic  $\text{O}^-$  ions. Most of the  $\text{O}_2^-$  ions have rather small kinetic energies, but few ions with kinetic energies of  $\approx 400\text{ eV}$  are also observed.

In addition, small amounts of negative  $\text{InO}^-$  ions were detected. The total intensity is about two orders of magnitude smaller than compared to the  $\text{O}^-$  ions. As shown in Figure 13, a significant fraction of  $\text{InO}^-$  ions has large kinetic energies. The energy distribution is somewhat different compared to  $\text{O}^-$  ions, however, suggesting a more complicated formation processes, however.

### 3.7 Energy Influx

The energy influx to the substrate was measured for both metallic and oxidized targets. The influence of oxygen flow, substrate bias and oxidation state of the target was investigated. For a metallic target, the measured energy influx increases from  $10$  to  $11.5\text{ mJ cm}^{-2}\text{s}^{-1}$  with increasing oxygen flow (Fig. 14). For an oxidized target, the energy influx is about  $0.9\text{ mJ}$



**Fig. 13.** Kinetic energy distribution of negative oxygen and indium oxide ions ( $\circ$ ,  $O^-$ ;  $\bullet$ ,  $InO^-$ )

$\text{cm}^{-2}\text{s}^{-1}$  smaller compared to the metallic target; this difference appears to be independent of the oxygen flow.

In order to explain this behavior one has to consider the different contributions to the energy influx. Major contributions to the measured energy influx are due to sputtered neutral particles, electrons, positive and negative ions, and the released heat of condensation [13]. The contribution of sputtered particles to the energy influx is given by the product of flux density and mean kinetic energy. The energy distribution function of sputtered species was calculated with the help of a transport-of-ions-in-matter (TRIM) code [14]. The flux density was estimated from the deposition rate and the mass density (Figs. 9 and 10). Contributions due to positive ions and electrons were calculated with the help of Langmuir probe theory. Finally, the contribution due to the condensation of In, Sn and O particles has been determined from the measured deposition rate and taking the specific heat of condensation into account.

In case of a metallic target, no oxygen flow, and for floating potential conditions ( $U_{sub} \approx -11$  V) the calculation yields a total energy influx of  $7.3 \text{ mJ cm}^{-2}\text{s}^{-1}$ . The contribution of sputtered In/Sn atoms amounts to  $4.4 \text{ mJ cm}^{-2}\text{s}^{-1}$ ,  $0.2 \text{ mJ cm}^{-2}\text{s}^{-1}$  are due to the heat of condensation, while positive ions (mainly  $\text{Ar}^+$ ,  $\text{O}^+$ , and  $\text{O}_2^+$  ions) and electrons contribute  $2.0 \text{ mJ cm}^{-2}\text{s}^{-1}$  and  $0.7 \text{ mJ cm}^{-2}\text{s}^{-1}$ , respectively. The calculated energy influx is somewhat lower than the measured value of  $9.8 \text{ mJ cm}^{-2}\text{s}^{-1}$  (Fig. 14). The reason for the difference could be contributions of secondary electrons and backscattered Ar ions, which were not taken into account.

In case of an oxidized target and/or with oxygen flow additionally the influence of negative oxygen ions needs to be taken into account. With increasing oxygen flow the deposition rate decreases, which is only partly offset

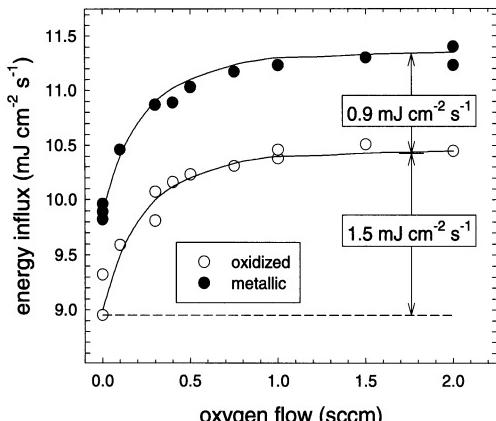


Fig. 14. Energy influx versus oxygen flows ( $U_{\text{sub}} = -11$  V)

by an increase in film density. Hence, we should expect a decrease of the energy influx with increasing oxygen flow, whereas an increase is observed. The noted difference should be explained by negative ions of relatively high kinetic energy contributing to the energy influx to the substrate.

## 4 Summary

Results for plasma-assisted deposition of indium-tin-oxide films by reactive magnetron sputtering have been reported. Pronounced dependencies of, e.g., film composition, film density, and deposition rate on oxygen flow, substrate bias voltage, and the oxidation state of the target (metallic or oxidized) have been observed. The observed dependencies are partly attributed to the changing energy-influx to the substrate and to the role played by positive and negative ions.

## References

1. K. L. Chopra, S. Major, D. K. Pandya, Thin Solid Films **102**, 1 (1983).
2. P. K. Song, Y. Shigesato, I. Yasui, C. W. Ow-Yang, D. C. Paine, Jap. J. Appl. Phys. **37**, 1870 (1998).
3. M. Bender, J. Trube, J. Stollenwerk, Appl. Phys. A **69**, 397 (1999).
4. L. J. Meng, M. P. dos Santos, Appl. Surf. Sci. **120**, 243 (1997).
5. D. Mergel, W. Stass, G. Ehl, D. Barthel, J. Appl. Phys. **88**, 2437 (2000).
6. M. Quaas, H. Steffen, R. Hippler, H. Wulff, Surf. Sci. **454**, 790 (2000).
7. H. Wulff, M. Quaas, H. Steffen, Thin Solid Films **355**, 395 (1999).
8. H. Wulff, H. Steffen, In: *Low Temperature Plasma Physics* (R. Hippler, S. Pfau, M. Schmidt, K. H. Schoenbach, eds.), Berlin: Wiley-VCH, p. 253 (2001), and references therein.
9. H. Wulff, M. Quaas, H. Steffen, R. Hippler, Thin Solid Films **377-378**, 418 (2000).

10. H. Kersten, E. Stoffels, W. W. Stoffels, M. Otte, C. Csambal, H. Deutsch, R. Hippler, *J. Appl. Phys.* **87**, 3637 (2000).
11. S. M. Rossnagel, J. J. Cuomo, W. D. Westwood (eds.) *Handbook of Plasma Processing Technology*. NOYES Publications: New Jersey, p. 234 (1990).
12. JCPDS-ICDD 5-642.
13. H. Kersten, H. Deutsch, H. Steffen, G. M. W. Kroesen, R. Hippler, *Vacuum* **63**, 385 (2001).
14. J. F. Ziegler, J. P. Biersack, U. Littmark, *The Stopping and Range of Ions in Solids*, Pergamon Press, New York (2003).

# Strains and Stresses in GaN Heteroepitaxy – Sources and Control

A. Dadgar<sup>1</sup>, R. Clos<sup>1</sup>, G. Strassburger<sup>1</sup>, F. Schulze<sup>1</sup>, P. Veit<sup>1</sup>, T. Hempel<sup>1</sup>, J. Bläsing<sup>1</sup>, A. Krtschil<sup>1</sup>, I. Daumiller<sup>2</sup>, M. Kunze<sup>2</sup>, A. Kaluza<sup>3</sup>, A. Modlich<sup>3</sup>, M. Kamp<sup>3</sup>, A. Diez<sup>1</sup>, J. Christen<sup>1</sup>, and A. Krost<sup>1</sup>

<sup>1</sup> Otto-von-Guericke-Universität Magdeburg  
FNW-IEP, Universitätsplatz 2, 39106 Magdeburg, Germany

<sup>2</sup> MicroGaN GmbH  
Albert-Einstein-Allee 45, 89081 Ulm, Germany

<sup>3</sup> Global Light Industries GmbH  
Carl-Friedrich-Gauß-Str. 1, 47475 Kamp-Lintfort, Germany

**Abstract.** We present a study of the sources of strain in GaN heteroepitaxy by in- and ex-situ measurement techniques. With an in-situ curvature measurement technique the strain development can be directly correlated to the different layers and doping in simple and device structures. We show several solutions for strain reduction and control. High-quality devices grown on Si are demonstrated.

Despite its great commercial success for the mass production of blue and green light-emitting diodes the epitaxial growth of GaN is still hindered by the lack of homosubstrates. Thus epitaxial growth is commonly performed on heterosubstrates as sapphire, SiC or Si. These substrates all have a large lattice (3.5–17%) and thermal (-25–116%) mismatch leading to difficulties in heteroepitaxial growth and are a high number of threading dislocations for all substrates and cracking for layers exceeding 1 and 3  $\mu\text{m}$  on Si and SiC, respectively. Especially the latter problem hindered GaN device layer growth on Silicon until the late 90s. After first reports on cracked, but functioning LED structures on Si [1,2] grown by molecular beam epitaxy (MBE) several methods to reduce cracking and devices were presented [3,4,5,6,7]. It turned out that strain engineering was crucial for the successful growth of device structures on Si. Here, we present a study on the sources of strains and stresses in GaN heteroepitaxy, which is not only valid for the growth on silicon, but also on other heterosubstrates as sapphire and SiC.

## 1 Experimental

For this study different GaN layer structures were grown by metal-organic chemical vapor phase epitaxy (MOVPE) on Si(111) substrates. Details on the growth process can be found elsewhere [8]. To directly study stress during epitaxial growth we applied an in-situ curvature measurement technique which is described in Ref. [9]. The stress can be calculated from the curvature using

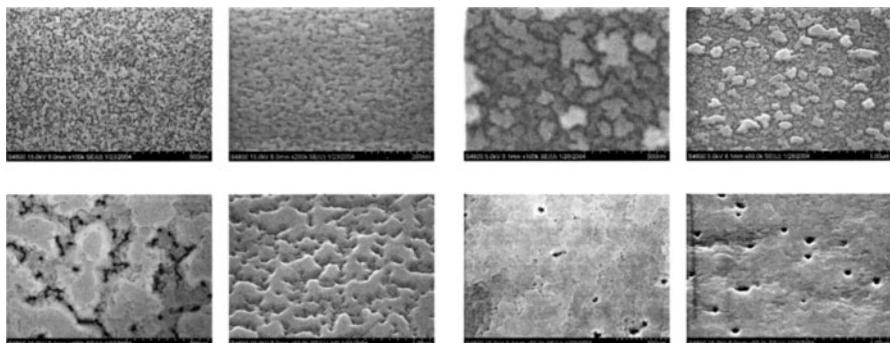
Stoney's equation [10] which, however, is only valid for low curvatures. In our case the curvatures often are too large. Therefore, we developed an analytical solution which gives an exact measure of stress [11] and which can be applied in the case of stronger curvatures. Additionally, for some samples, ex-situ measurements as x-ray diffraction (XRD), transmission electron microscopy (TEM), scanning electron microscopy (SEM), and photoluminescence (PL) were performed.

## 2 Sources of Stresses and Strains

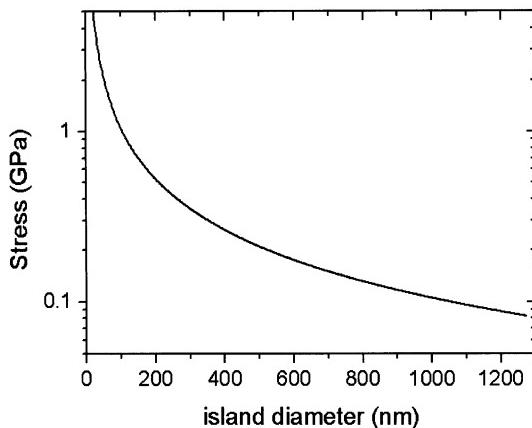
### 2.1 Seed Layer Growth

The heteroepitaxial growth usually starts with a so-called seed layer which helps to overcome the wetting problems on the heterosubstrate originating in the differences of the surface energies. In the case of Si and SiC, this is usually a thin AlN layer deposited at high or low growth temperatures. The typical structure of such a layer is columnar with grain diameters between 20–40 nm [12]. Subsequent GaN growth then leads to the formation of islands of different size some of which grow faster and overgrow smaller islands as shown in the SEM images in Fig. 1.

Before islands coalesce the gap between these independently grown crystallites has to be closed. But in this final step an additional Ga-N molecule (or a monolayer) usually does not exactly fit into the gap. Therefore, when the gap is smaller than a lattice distance (e.g. two times or  $\sqrt{3}$  times the lattice parameter  $a$ , for the  $<\bar{1}2\bar{1}0>$  and  $<10\bar{1}0>$  directions, respectively), the gap can be closed by introducing additional Ga-N molecules inducing a local compressive strain or by closing the gap without the addition of Ga-N molecules which will induce a local tensile strain into the lattice. The strain

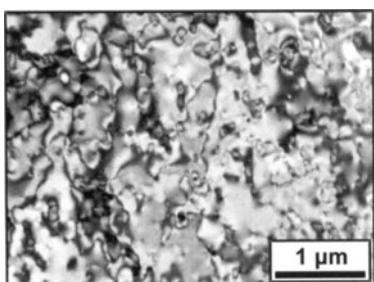


**Fig. 1.** SEM images of the growth of GaN on a 25 nm thick AlN seed layer on Si. Images show plan-view (left) and birds-eye-view (right) images after 0, 15, 60 and 120 s of GaN growth with nominal GaN thicknesses of 0, 10, 42, and 84 nm, respectively (from upper left to lower right)

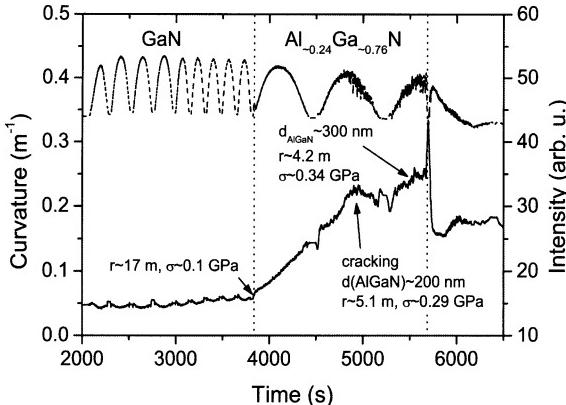


**Fig. 2.** Stress induced by island coalescence calculated for hexagonal islands

can deform the following planes and facilitate or hinder the further closure of the gap for tensile or compressive strain, respectively. Since we do only observe tensely strained layers, the second mechanism is dominant. Consequently, the amount of such boundaries, which is correlated to the density of crystallites, determines the amount of tensile strain in the layer. Similar observations were made for the growth of GaN on sapphire [13]. Based on these assumptions a simple calculation shows that the tensile stress for a grain diameter of 100 nm is about 1 GPa, for 300 nm it is only 0.35 GPa (Fig. 2). Typical crystallite diameters range from 200–400 nm for a 500 nm thick GaN layer on an AlN-seed on Si. In plan-view TEM images one can see strong contrast fluctuations likely originating from local tensile stresses at island boundaries (Fig. 3). The boundaries of these crystallites are also the main source of threading dislocations; consequently, larger crystallites lead to a decrease in the dislocation density.



**Fig. 3.** Plan-view TEM image of a GaN layer on Si. Local strain originating in gap closure induces a bending of the thin GaN foil leading to contrast fluctuations



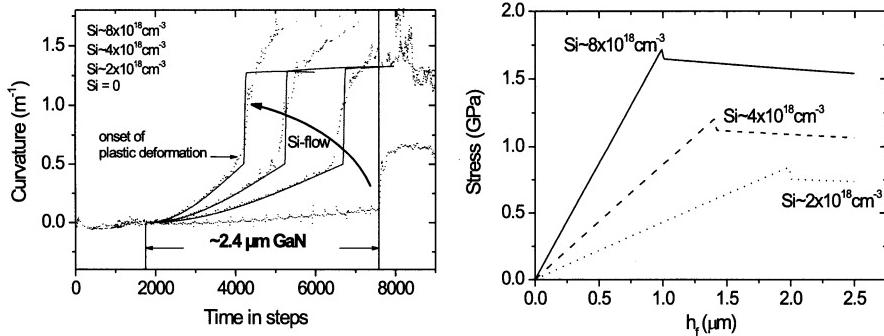
**Fig. 4.** Curvature (lower curve) and reflected intensity (upper curve) of the growth of  $\sim 350$  nm  $\text{Al}_{\sim 0.24}\text{Ga}_{\sim 0.76}\text{N}$  on  $\sim 1750$  nm GaN. Tensile stress leads to cracking and partial relaxation during AlGaN growth

## 2.2 Lattice Mismatch

Ternary alloys, e.g. for device layers as FETs, LEDs, or lasers introduce strong strain in the GaN system. In most material systems a strong lattice mismatch relaxes by the formation of dislocations, but in the case of the group-III-nitrides, and especially for AlN or AlGaN with high Al content, cracking is also observed. To avoid the formation of cracks and dislocations lattice mismatched layers are usually grown in thicknesses below the critical thickness. Then they grow pseudomorphically with the lattice parameter of the underlying buffer layer. Figure 4 shows an in-situ curvature measurement of the growth of  $\text{Al}_{\sim 0.24}\text{Ga}_{\sim 0.76}\text{N}$  on GaN. The lattice mismatch of  $\Delta a/a \sim 0.006$  leads to a strong tensile stress, and consequently, a strong wafer bowing is observed. Interestingly, one can even observe relaxation by cracking during growth. In our case this relaxation appears the first time after  $\sim 200$  nm of AlGaN growth. At this point, additionally to the 0.1 GPa stress of the 1750 nm thick GaN layer, 0.19 GPa are accumulated from the lattice mismatched AlGaN growth. The relaxation can be directly seen in the change of the radius of curvature and indirectly by the increase in scattering of the reflected intensity which is influenced by the cracks (Fig. 4). In the case of AlN growth the growth temperature determines whether AlN grows pseudomorphic or partly relaxed. We find a transition from pseudomorphic to relaxed growth around  $1000 - 1100$  °C [9] which was also observed in studies of AlN seed layer growth by other authors [14,15]. This can be used for strain engineering as will be shown later.

## 2.3 Doping

To achieve n-type conductivity in GaN it is commonly doped with Si. When doped with high Si concentrations above  $10^{18} \text{ cm}^{-3}$  cracking of the GaN layer

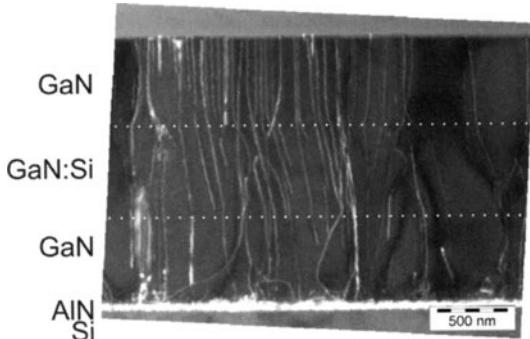


**Fig. 5.** Curvature and fit of differently high Si-doped GaN layers during MOVPE growth (left) and stress (right) determined with a nonlinear analytical theory [11]. The high stress leads to a plastic deformation of the Si wafer

is sometimes observed [16,17]. From in-situ measurements we find a strong dependence of in-situ determined tensile stress with the amount of Si [9] as already reported by Romano et al. [16]. For the p-type dopant Mg it is significantly lower, although a higher doping concentration around  $5 \times 10^{19} \text{ cm}^{-3}$  is used. Here the Mg-induced tensile stress is comparable to a Si doping concentration around  $5 \times 10^{17} \text{ cm}^{-3}$ . Tensile stress originating from Si doping is significantly higher than intrinsic tensile stress from the 3D growth mode in the beginning and leads to such a strong wafer curvature that even plastic deformation of Si takes place (Fig. 5).

In TEM images a slight bending of dislocations is seen (Fig. 6). This can be also seen in the work of Romano et al. for edge type dislocations (see Fig. 3 in Ref. [16]) and even more pronounced at doping levels well above  $10^{19} \text{ cm}^{-3}$  for compressively grown AlGaN by Cantu et al. [18]. Most likely, Si, which is known as anti-surfactant in GaN epitaxy, decorates dislocations [19]. Thus for an edge-type dislocation Si blocks Ga from attaching at the dislocation core and the dislocation moves laterally and shortens the correlated lattice plane one unit cell. This model can explain the observations in TEM measurements (Fig. 7). In contrast to our findings Romanov and Speck [20] claim that dislocation bending is promoted by compressive stress as it is the case for their samples. However, we have an initially slightly tensely stressed layer with a further increasing tensile stress when the layer is Si doped which would not fit into their model. The explanation presented here, however would fit both observations since in both cases the system is getting more tensile.

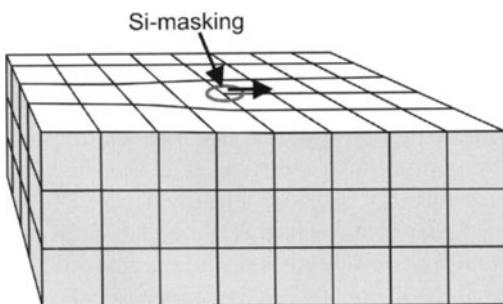
This movement always decreases the length of the dislocation-correlated lattice plane and by this induces tensile stress. Because no significant roughening of the surface is found in these doping experiments the dislocation movement is most likely due to a masking effect and not to a bending by a roughened surface. A bending by a roughened surface should also lead to a dislocation movement which would induce tensile and compressive stress which is not observed here.



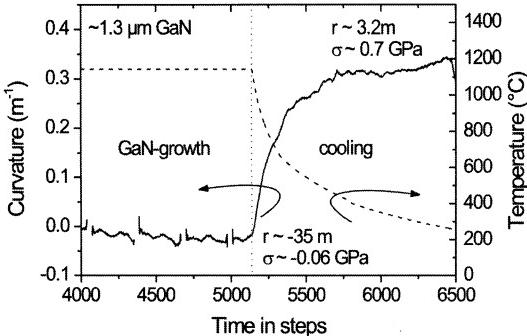
**Fig. 6.** Cross-section TEM image of a partially Si-doped AlN/GaN/GaN:Si/GaN layer on Si(111) as in Fig. 7. In this case Si doping leads to a dislocation bending of 10-13 °

## 2.4 Thermal Stress

GaN growth is usually performed at temperatures above 1000 ° C. Cooling introduces stress which is compressive for GaN on sapphire and tensile in the case of GaN on Si and SiC. From in-situ measurements we can determine stress during growth and after cooling (Fig. 8). Thus, we can subtract intrinsic or doping correlated stresses and determine thermal stress separately. It is found from several experiments with different layer thicknesses that cooling of GaN on Si induces a tensile stress of 0.6-0.7 GPa/ $\mu$ m. Finding ways to compensate tensile thermal stress is the key for the growth of high-quality GaN and GaN-based devices on Si. Even on SiC tensile thermal stress leads to cracks for layers in excess of 3  $\mu$ m.



**Fig. 7.** Illustration of the proposed mechanism leading to dislocation bending: Si masks the dislocation core and, acting as anti-surfactant, prevents further GaN growth at the dislocation core. Thus when the next layer is grown, the core moves one unit cell shortening the associated lattice plane



**Fig. 8.** In-situ curvature measurement at the end of 1.3  $\mu\text{m}$  GaN growth and subsequent cooling. From the difference in wafer curvature we determine a tensile thermal stress of  $\sim 0.6 \text{ GPa}/\mu\text{m}$

### 3 Strain Engineering

#### 3.1 Thermal Stress

Thermal tensile stress can be relieved in two principally different ways. One is the growth on fields small enough that cracks won't occur. This would mean that with a crack distance of 50  $\mu\text{m}$  and below one could only grow miniaturized devices. But with strain-relieving buffer layers [21] device-relevant thicknesses and field sizes have been demonstrated [22,23,24,25]. The other approach is to compensate tensile stress by a sufficiently large compressive stress [3,4,7]. One of the simplest and most universal approaches is to use relaxed-grown thin AlN interlayers [6]. AlN has a tendency to grow polycrystalline below 1000–1100 °C [14,15]. Thus, it grows fully or partially relaxed on GaN [9]. The subsequently grown GaN will grow compressively strained on top of them compensating tensile thermal stress. Figure 9 shows the in-situ curvature measurement of two identically thick structures with an approximately 12 nm thick high-temperature (HT) and a low-temperature (LT) AlN interlayer grown at GaN growth temperature and 630 °C, respectively. Properly adjusting LT-AlN interlayers helps to grow GaN layers which are unstrained after growth. This is important for device processing where a low wafer curvature is prerequisite. With this method up to 7  $\mu\text{m}$  thick layers with multiple AlN interlayers were already demonstrated [9].

#### 3.2 Intrinsic Stress

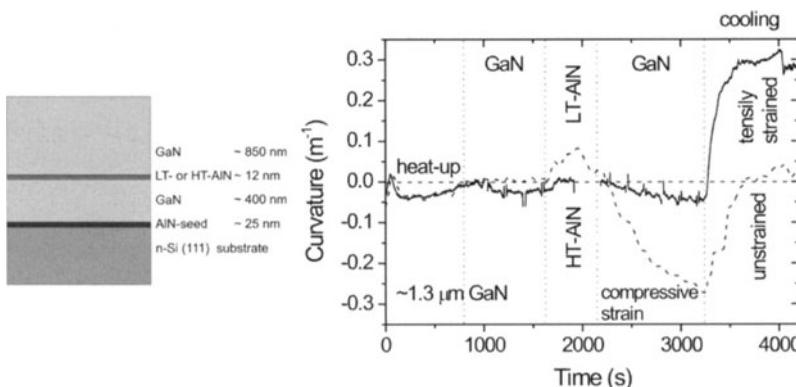
To reduce intrinsic stress a reduction of grain boundaries is necessary. Thus, larger grains have to be grown. On sapphire this is relatively simple by using a GaN seed layer which recrystallizes during heat up and forms few islands on the surface from which high-temperature growth will proceed [13]. Using an AlN seed layer this is not possible because it always covers the surface

and due to its high binding energy cannot be reorganized as GaN by a heat treatment.

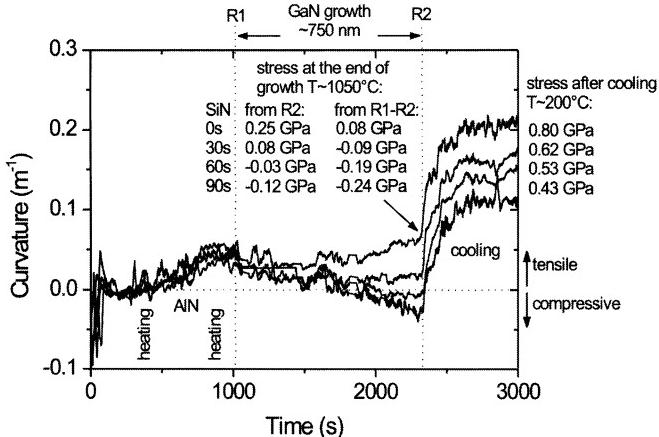
A simple way to grow larger grains is to mask part of the seed layer with a SiN layer in the nominal range of one to two monolayers [26]. Since SiN acts as an anti-surfactant [19] GaN will only grow in uncovered regions and larger grains can develop reducing the amount of boundaries and by this reducing the tensile stress. An additional positive effect of initially larger GaN crystallites is the reduction in the dislocation density [26,27]. Figure 10 shows an in-situ curvature measurement of stress development for differently long SiN deposition times on the AlN seed layer. Due to the increase in island size with increasing SiN deposition time GaN grows under decreasing tensile stress. Interestingly, even a slight compressive stress is observed for long SiN deposition times. This total compressive stress can be explained by a hydrostatic stress of the islands originating in the surface energy and only a small portion of tensile stress from island coalescence.

### 3.3 Doping

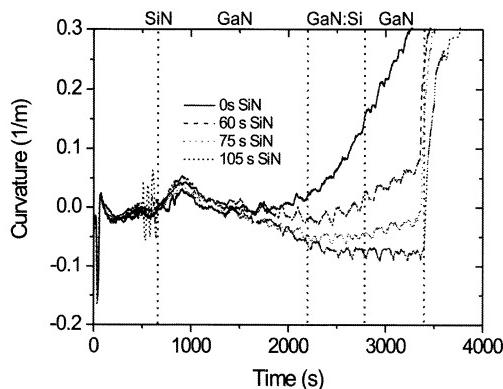
For many devices as LEDs and lasers high doping concentrations and good carrier mobilities are requested. Especially, for large-area LEDs on sapphire a good n-type back contact requires good lateral current spreading. Thus, already the buffer layer which contains a high number of dislocations must be doped with high Si concentrations. As described previously the masking of dislocations and the reduction of the length of the correlated lattice plane might be the strain generating mechanism. Therefore, tensile stress can be best avoided by reducing the amount of dislocations. One way to achieve this is to use thin SiN masking layers which are helpful to reduce the dislocation density by a factor of 10 and more. Already one order of magnitude less dislocations leads to a reduction in tensile stress to a tenth which is sufficient



**Fig. 9.** Curvature measurement (right) of two identically thick GaN on Si samples (left: sample scheme) containing a high- or a low-temperature AlN interlayer



**Fig. 10.** Stress development of a GaN layer grown on an AlN seed layer with differently thick SiN masks. SiN deposition times range from 0–90 s

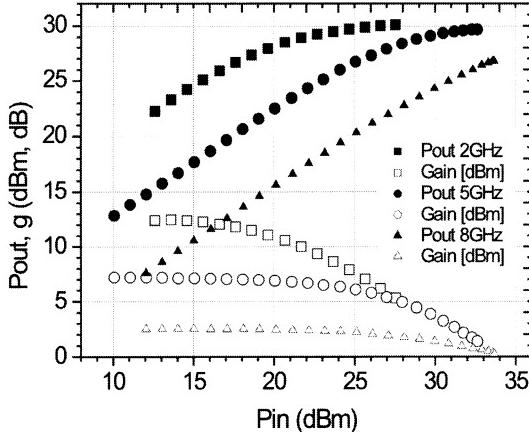


**Fig. 11.** Stress development during the growth of  $8 \times 10^{18} \text{ cm}^{-3}$  Si-doped GaN on buffer layers with different dislocation densities achieved by differently thick SiN in-situ masks in the buffer layer

in most cases. When growing GaN layers with identical doping concentrations but a variation in the dislocation density [26] one observes a strong decrease in curvature which originates in a reduction of tensile stress (Fig. 11). This reduction in tensile stress can be attributed to a decrease in the dislocation density.

## 4 Devices

GaN is the base material for a variety of different device applications as LEDs, lasers, field-effect-transistors (FETs) and for sensor devices as, e.g. ph- or gas-sensors [28]. The problem for the fabrication of high-quality devices is



**Fig. 12.** Power measurements of an AlGaN/GaN FET on Si at different frequencies

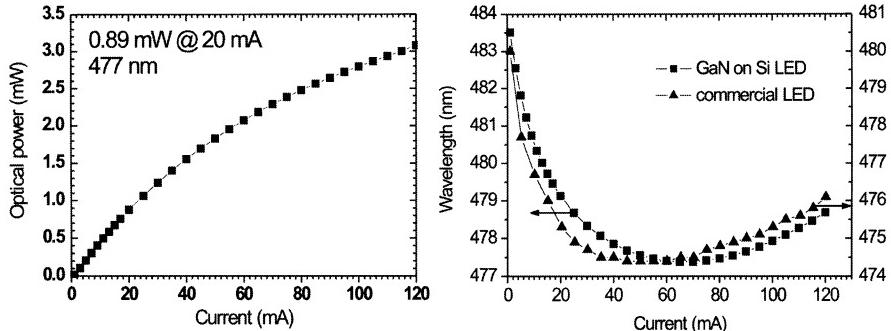
that thick layers are difficult to achieve on Si. But thick, high-quality layers are prerequisite for devices as LEDs which can be only achieved by strain engineering.

#### 4.1 Transistors

Simple 1.3-1.8  $\mu\text{m}$  thick AlGaN/GaN transistor structures were grown on Si(111) substrates with an AlN-seed layer, a GaN buffer of  $\sim 400$  nm, a LT-AlN interlayer, and  $> 800$  nm GaN. The active layer consists of 25–28 nm thick  $\text{Al}_{\sim 0.25}\text{Ga}_{\sim 0.75}\text{N}$ . With such structures we can routinely achieve room temperature carrier mobilities above  $1500 \text{ cm}^2/\text{Vs}$  at sheet carrier concentrations of  $1\text{--}1.3 \times 10^{13} \text{ cm}^{-3}$ . A major problem for high-frequency operation are the power losses due to damping in the highly-resistive but, compared to intrinsic GaN, still conductive Si substrate. Here, a major problem is that all elements used for GaN and AlGaN epitaxy can act as dopants of the Si substrate [29]. Additionally, the growth temperatures can lead to an increase in the carrier concentration of the highly-resistive undoped Si substrate. We have achieved best devices with an output power of 2.2 and 2 W/mm at 2 and 5 GHz, respectively which reduces to 0.9 W/mm at 8 GHz due to absorption losses in the Si substrate (Fig. 12).

#### 4.2 LEDs

The first GaN LED on Si presented by Guha et al. [1,2] was a MBE-grown thin and rather dim device full of cracks. Prerequisite to achieve high output powers are low dislocation densities below  $10^9 \text{ cm}^{-2}$ . This requires the growth of thick buffer layers to enable dislocation recombination. But thick crack-free layers are difficult to grow on Si and there are only few reports on crack-free



**Fig. 13.** Power vs. current (left) of a packaged InGaN LED on Si. For comparison of the wavelength shift vs. current (right) a packaged InGaN LED on Si and a commercial InGaN LED on sapphire are shown. A blue shift due to band filling up to currents of 70 mA (55 mA for the LED on sapphire) followed by a red shift due to heating effects for higher currents is observed. The wavelength shift data shows a slightly better behavior than the commercial device on sapphire, likely due to the higher heat conductivity of the silicon substrate

LEDs in excess of 2  $\mu\text{m}$ . With LT-AlN interlayers for strain compensation and thin SiN masking layers we are able to reduce the dislocation density to  $10^9 \text{ cm}^{-2}$  [23] and achieve thicknesses up to 3.7  $\mu\text{m}$  with a low radius of curvature around 20  $\mu\text{m}$  of the device layers at room temperature, important for further processing.

Best results obtained so far are output powers of 0.89 mW at 20 mA and 477 nm (Fig. 13). These packaged devices can be driven up to 120 mA cw with an output power of >3 mW without degradation. No degradation or device failure has been observed in continuous operation of 40 LEDs at 30 mA for half a year.

## 5 Summary

We have shown different sources of strain in GaN heteroepitaxy by in-situ stress measurements. Most of these stress sources can be compensated by growing appropriate layer schemes. With such improved layers we have grown thick device structures with excellent performance data for devices on Si.

## Acknowledgements

We like to acknowledge the Deutsche Forschungsgemeinschaft for their support in the framework of contract numbers KR1239/10-2 and Ch87/4-1.

## References

1. S. Guha and N. A. Bojarczuk, Appl. Phys. Lett. **72**, 415 (1998).

2. S. Guha and N. A. Bojarczuk, Appl. Phys. Lett. **73**, 1487 (1998).
3. E. Feltin, S. Dalmasso, P. de Mierry, B. Beaumont, H. Lahrèche, A. Bouillé, H. Haas, M. Leroux, and P. Gibart, Jpn. J. Appl. Phys. **40**, L738 (2001).
4. H. Marchand, L. Zhao, N. Zhang, B. Moran, R. Coffie, U.K. Mishra, J.S. Speck, S.P. DenBaars, and J.A. Freitas, J. Appl. Phys. **89**, 7846(2001).
5. A. Dadgar, A. Alam, T. Riemann, J. Bläsing, A. Diez, M. Poschenrieder, M. Strassburg, M. Heuken, J. Christen und A. Krost, Phys. Stat. Sol. (a) **188**, 155 (2001).
6. Armin Dadgar, Jürgen Bläsing, Annette Diez, Assadullah Alam, Michael Heuken und Alois Krost, Jpn. J. Appl. Phys. **39**, L1183 (2000).
7. A. Vescan, J.D. Brown, J.W. Johnson, R. Therrien, T. Gehrke, P. Rajagopal, J.C. Roberts, S. Singhal, W. Nagy, R. Borges, E. Piner, K. Linthicum, Phys. Stat. Sol. (c) **0**, 52 (2003).
8. Armin Dadgar and Alois Krost: MOVPE growth of GaN on Si, in *Vacuum Science and Technology: Nitrides as seen by the Technology, 2002*, Tanya Paskova and Bo Monemar (Eds.), (Research Signpost, Kerala (India) 2002).
9. A. Krost, A. Dadgar, G. Strassburger, and R. Clos, Phys. Stat. Sol. (a) **200**, 26 (2003).
10. G. Stoney, Proc. R. Soc. Lond. A **82**, 172 (1909).
11. R. Clos, A. Dadgar, and A. Krost, to be published.
12. D.M. Follstaedt, J. Han, P. Provencio, and J.G. Fleming, MRS Internet J. Nitride Semicond. Res. **4S1**, G3.72 (1999).
13. T. Böttcher, S. Einfeldt, S. Figge, R. Chierchia, H. Heinke, D. Hommel, and J. S. Speck, Appl. Phys. Lett. **78**, 1976 (2001).
14. A. Watanabe, T. Takeuchi, K. Hirosawa, H. Amano, K. Hiramatsu, and I. Akasaki, J. Cryst. Growth **128**, 391(1993).
15. H. Lahrèche, P. Vennégùès, O. Totterau, M. Laügt, P. Lorenzini, M. Leroux, B. Beaumont, and P. Gibart, J. Cryst. Growth **217**, 13 (2000).
16. L. T. Romano, C. G. Van de Walle, J. W. Ager III, W. Götz, and R. S. Kern, J. Appl. Phys. **87**, 7745 (2000).
17. E. K. Sia, M. S. Hao, Soo-Jin Chua, I. M. Tiginyanu, V. Ichizli, Kabula Mutambwa, Hans L. Hartnagel, Ji Zhang, and S. Tripathy, Proc. SPIE Int. Soc. Opt. Eng. **4594**, 201 (2001).
18. P. Cantu, F. Wu, P. Waltereit, S. Keller, A. E. Romanov, U. K. Mishra, S. P. DenBaars, and J. S. Speck Appl. Phys. Lett. **83**, 674 (2003).
19. Satoru Tanaka, Misaichi Takeuchi and Yoshinobu Aoyagi, Jpn. J. Appl. Phys. **39**, L1249(2000).
20. A. E. Romanov and J. S. Speck, Appl. Phys. Lett. **83**, 2569 (2003).
21. A. Dadgar, A. Alam, J. Christen, T. Riemann, S. Richter, J. Bläsing, A. Diez, M. Heuken, and A. Krost, Appl. Phys. Lett. **78**, 2211 (2001).q
22. A. Krost and A. Dadgar, Mat. Sci. Eng. **B93**, 77 (2002).
23. A. Dadgar, M. Poschenrieder, O. Contreras, J. Christen, K. Fehse, J. Bläsing, A. Diez, F. Schulze, T. Riemann, F.A. Ponce, and A. Krost, Phys. Stat. Sol. (a) **192**, 308 (2002).
24. A. Dadgar, M. Poschenrieder, J. Bläsing, K. Fehse, A. Diez, and A. Krost, Appl. Phys. Lett. **80**, 3670 (2002).
25. Y. Honda, Y. Kuroiwa, M. Yamaguchi, and N. Sawaki, Appl. Phys. Lett. **80**, 222 (2002).
26. A. Dadgar, M. Poschenrieder, A. Reiher, J. Bläsing, J. Christen, A. Krtschil, T. Finger, T. Hempel, A. Diez, and A. Krost, Appl. Phys. Lett. **82**, 28 (2003).

27. O. Contreras, F.A. Ponce, J. Christen, A. Dadgar, and A. Krost, *Appl. Phys. Lett.* **81**, 4712 (2002).
28. M. Eickhoff, J. Schalwig, G. Steinhoff, O. Weidemann, L. Görgens, R. Neuberger, M. Hermann, B. Baur, G. Müller, O. Ambacher, M. Stutzmann, *Phys. Stat. Sol. (c)* **0**, 1908 (2003).
29. Pradeep Rajagopal, John C. Roberts, J. W. Cook, Jr., J. D. Brown, Thomas Gehrke, Edwin L. Piner, and Kevin J. Linthicum, Proc. of MRS Fall Meeting 2003 Boston USA.

# Morphological Transition in Thin Lamellar Diblock Copolymer Films as Revealed by Combined GISAXS and AFM Studies

Christine M. Papadakis<sup>1\*</sup>, Peter Busch<sup>1\*\*</sup>,  
Dorthe Posselt<sup>2</sup>, and Detlef-M. Smilgies<sup>3</sup>

<sup>1</sup> Fakultät für Physik und Geowissenschaften, Universität Leipzig  
Linnéstr. 5, D-04103 Leipzig, Germany

<sup>2</sup> IMFUFA (Department of Mathematics and Physics), Roskilde University  
P.O. Box 260, DK-4000 Roskilde, Denmark

<sup>3</sup> Cornell High-Energy Synchrotron Source (CHESS) G-line Division  
Cornell University, Ithaca, NY 14853, USA

**Abstract.** Nanostructured polymer films are of increasing importance for many applications. A number of experimental methods have been used to investigate their structures, e.g., imaging in real space and scattering methods using x-rays or neutrons. We briefly review these methods and describe results from atomic force microscopy and grazing-incidence small-angle x-ray scattering measurements on thin films of lamellar poly(styrene-*b*-butadiene) (P(S-*b*-B)) diblock copolymer melts on Si substrates with a native oxide layer. For high molar masses, P(S-*b*-B) spontaneously forms structured surfaces due to a perpendicular orientation of the lamellae with respect to the film surface. The results from P(S-*b*-B) are compared to the frequently studied poly(styrene-*b*-methylmethacrylate) (P(S-*b*-MMA)) system. The findings can be understood by comparing the enthalpic and entropic contributions to the free energy of the copolymers at the film surface and the film/substrate interface.

## 1 Introduction

Nanostructured surfaces formed by diblock copolymer films are of increasing importance for many different purposes, as described in two recent reviews [1,2]. The attractive feature of block copolymers is their spontaneous tendency to self-organize into ordered structures (e.g., lamellae, cylinders on a hexagonal lattice, spheres on a body-centered cubic lattice) which can be controlled by choosing the block copolymer molar mass and the relative block volumes. The pattern repeat distance is typically of the order of 100–1000 Å, thus making nanolithography possible. By suitably modifying one of the blocks, selective etching allow the mesoscopic structures formed by poly(styrene-*b*-polybutadiene) (P(S-*b*-B)) or poly(styrene-*b*-polyisoprene)

---

\* New address: Physik department E13, TU München, James-Franck Str. 1, 85747 Garching, [christine.papadakis@ph.tum.de](mailto:christine.papadakis@ph.tum.de)

\*\* New address: Institut für Polymerforschung e.V., Hohe Strae 6, 01069 Dresden

(P(S-*b*-I)) films to be exploited in generating periodically arranged holes and dots with a density of  $\sim 10^{11}$  holes /cm<sup>2</sup> [3] in the substrate. Similarly three-dimensionally structured surfaces can be achieved by making use of the wettability contrast with respect to mica of the two blocks in micelle-forming poly(styrene-*b*-2-vinylpyridine) (P(S-*b*-2VP)) diblock copolymers [4]. Using P(S-*b*-2VP) as a template, it is also possible to produce gold nanoclusters of a few nanometers size with distances of up to micrometers as well as nanowires of a few nanometers in width and of up to 50  $\mu\text{m}$  in length, both attached to the substrate [5].

The cylindrical structure formed by poly(styrene-*b*-methylmethacrylate) (P(S-*b*-MMA)) diblock copolymers can be used for creating nanowire arrays [6,7]. Application of an electrical field perpendicular to the substrate leads to dense arrays of cylinders perpendicular to the substrate. The cylinders can be selectively removed by deep ultraviolet exposure in order to fill them with ferromagnetic cobalt [7].

Prepatterned substrates have been used to control the orientation of the mesoscopic block copolymer structures in thin film geometry and in this way to obtain structured surfaces with macroscopic order [8,9]. Furthermore, applying outer fields may orient the structures during or after film preparation [6,10,11].

Thin, lamellar films based on diblock copolymers may be of interest for a variety of applications, for instance as photonic band gap materials [12], as wave guides and for the confinement of particles within the lamellar layers [14]. Additionally, films with the lamellae parallel to the film surface offer the possibility to switch the surface properties by changing the environment making the buried block migrate to the film surface.

An understanding of the mechanisms governing the structurization of thin, supported diblock copolymer films is of key importance for a controllable preparation of such structures. A number of experimental methods have been applied in this context, and we briefly review these before describing our results on lamellar P(S-*b*-B) thin films on Si substrates with a native SiO<sub>x</sub> layer (denoted Si/SiO<sub>x</sub> in the following) obtained using atomic force microscopy (AFM) and grazing-incidence small-angle X-ray scattering (GISAXS, [14,15,16,17]). The results are compared to the findings reported in the literature on supported thin films of lamellar P(S-*b*-MMA) on Si/SiO<sub>x</sub>. The main differences between these two systems are discussed in terms of the enthalpic and entropic contributions to the free energy of the copolymers at the film surface and the film/substrate interface.

## 2 Methods Used for Investigating Thin Lamellar Diblock Copolymer Films

Numerous experimental methods have been used to investigate the structure of thin diblock copolymer films, among these both imaging methods, provid-

ing information in real space, and scattering methods, providing information in reciprocal space [18,19,20,21].

Optical microscopy has been used to image the surface morphology of thin diblock copolymer films. The bimodal height histograms obtained with lamellar diblock copolymer films indicate for certain systems terraces at the film surface with terrace heights equal to the bulk lamellar thickness,  $D_{\text{lam}}$ , i.e., the lamellae are parallel to the film surface, and islands or holes form in the uppermost layer due to an incommensurability between the film thickness,  $D_{\text{film}}$  and  $D_{\text{lam}}$  [22,23,24,25,26].

Atomic force microscopy (AFM) offers a much better spatial resolution (usually a few nm), and the details of the terrace edges and the shapes of the islands in lamellar diblock copolymer films with the lamellae parallel to the film surface could be resolved [26,27,28,29]. Also, in case of structured surfaces, the block copolymer mesostructure at the film surface can be made visible, and the local details can be studied [8,17,30,31]. Usually, AFM measurements of block copolymer films are conducted in tapping mode, i.e., with the tip oscillating at its resonance frequency barely touching the surface, in order to avoid damaging the soft polymer films. This way, in addition to the height images recorded at constant damping, phase images (variations of the phase shift of the oscillating cantilever during scanning) can be obtained which are related to the material contrast. This is particularly interesting when the viscoelastic properties of the two blocks are significantly different, as is the case in P(S-*b*-B), for instance [17].

Another imaging technique frequently used is electron microscopy either in plan view [3,32] or in cross-section geometry [33,34,35]. Usually, the contrast is achieved by selectively staining one of the blocks by a heavy metal, e.g., by exposing the samples to OsO<sub>4</sub> vapor. Microtoming and sample handling may be difficult, especially for samples with glass transition temperatures below room temperature.

Imaging techniques have the advantage of revealing details of the local film structure, such as defects; however for obtaining depth-dependent information, the films need to be destroyed. Scattering techniques do not suffer from this drawback, and, in addition, accurately statistically averaged information averaged over the sample is obtained. Small-angle X-ray and neutron scattering in transmission geometry have been applied to study the mesoscopic structures of thin block copolymer films [36,37]. However, for satisfactory contrast in neutron scattering, partially deuterated polymers are mandatory for most systems. Furthermore, the (weak) scattering from films with typical thicknesses of a few hundred to a few 1000 Å can be obscured by scattering from the substrate (typical thickness 1 mm).

X-ray and neutron reflectometry map the reflectivity of the sample as a function of the incident angle,  $\alpha_i$ , and yield laterally averaged information on the electron density or the scattering length density, respectively, i.e., the concentration profile, along the film normal [18,21]. For a parallel orientation

of the lamellae in block copolymer films, the details of the lamellar structure can be revealed [38,39,40,41,42,43,44,45]. However, in the case of the perpendicular lamellar (or cylinder) orientation, no information about the inner film structure can be gained from reflectometry. Typical reflectivity measurements take at least an hour, and hence this method is not suitable for real-time observations.

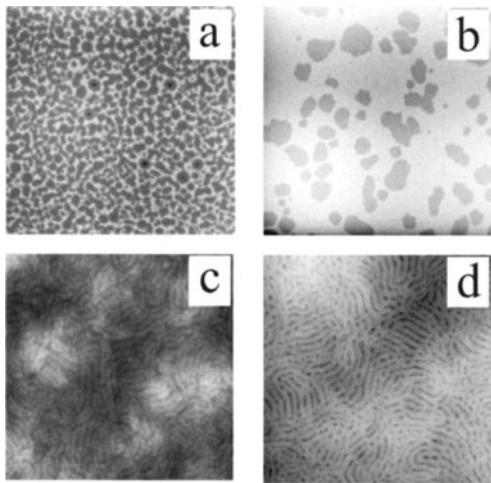
Grazing-incidence small-angle scattering with X-rays (GISAXS, [?,47]) and neutrons, (GISANS, [48,49]) have recently been applied to thin diblock copolymer films exploring the properties of the upper terraced layer in the parallel lamellar orientation [50] and, the dewetting of diblock copolymer films [48,49,51,52,53] as well as films displaying perpendicular lamellae [14].  $\alpha_i$  is chosen to be slightly above the critical angle of external reflection of the polymer film,  $\alpha_{cP}$ , which is typically of the order of  $10.1^\circ$ . Under these conditions, X-rays and neutrons penetrate films up to several  $1000\text{ \AA}$  completely. In this way, absorption by the substrate is minimized and the scattering volume maximized. The scattering in the plane of incidence, i.e., normal to the film surface ( $q_z$ ), and parallel to the film surface ( $q_y$ ) is recorded simultaneously by means of a two-dimensional detector. The GISAXS method has several advantages:

- In general, no special sample preparation is in general necessary.
- It is non-destructive.
- As a large area (a few  $\text{mm}^2$ ) is illuminated, accurately statistically averaged information about the sample is obtained.
- Under certain conditions, measuring times as low as seconds lead to satisfying scattering images, making time-resolved measurements possible.

However, the data analysis in GISAXS requires some caution, especially at small incident angles, since it is necessary to take into account that the incident beam is not only scattered by the polymer film, but it may also be reflected by the film/substrate interface before and/or after the scattering event, leading to additional intensity. In addition, the refraction of the beam at the film surface has to be considered [54]. These dynamical effects may be accounted for in the analysis by the distorted-wave Born approximation (DWBA, [55,56,57,58,59]). Only for  $\alpha_i$  sufficiently larger than  $\alpha_{cP}$ , the Born approximation (single scattering only) is recovered.

### 3 The Lamellar Orientation in Thin Poly(styrene-*b*-butadiene) Films

We have recently reported the lamellar orientation in thin, supported films of P(S-*b*-B) diblock copolymers as found using AFM [17]. The samples were spin-coated from toluene solution onto Si/SiO<sub>x</sub>. The film thickness was controlled by casting films from solutions of different polymer concentrations (0.1–7 wt.-%). For details of the preparation, see [17]. Typical images from

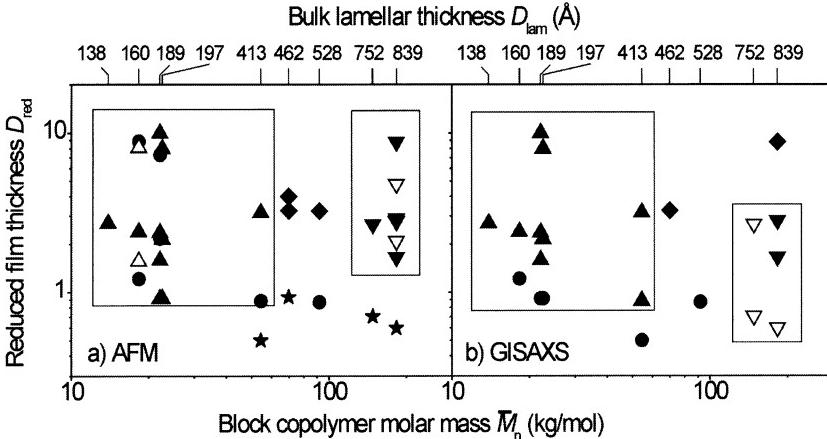


**Fig. 1.** AFM images from two low molar mass ( $\bar{M}_n = 22.6 \text{ kg/mol}$ ,  $D_{\text{lam}} = 197 \text{ \AA}$ , upper row) and two high molar mass P(S-*b*-B) films ( $183 \text{ kg/mol}$ ,  $D_{\text{lam}} = 839 \text{ \AA}$ , lower row) with different reduced film thicknesses [17]: (a)  $D_{\text{red}} = 2.1$ , image size  $50 \times 50 \mu\text{m}^2$ , height scale  $500 \text{ \AA}$ , (b)  $D_{\text{red}} = 8.0$ , image size  $100 \times 100 \mu\text{m}^2$ , height scale  $1200 \text{ \AA}$ , (c)  $D_{\text{red}} = 2.8$ , image size  $3 \times 3 \mu\text{m}^2$ , height scale  $200 \text{ \AA}$ , (d)  $D_{\text{red}} = 8.8$ , image size  $3 \times 3 \mu\text{m}^2$ , height scale  $400 \text{ \AA}$ . In all images, the height scale ranges from black (low) to white (high)

tapping mode AFM on a low and a high molar mass sample for low and high values of the reduced film thickness,  $D_{\text{red}} = D_{\text{film}}/D_{\text{lam}}$  are compiled in Fig. 1. The low molar mass films (upper row) display holes in the surface layer with heights similar to the bulk lamellar thickness,  $D_{\text{lam}}$ , regardless of  $D_{\text{red}}$ . The surfaces of the high molar mass films (lower row), on the other hand, display a lamellar texture. For  $D_{\text{red}} = 2.8$ , the repeat distance ( $855 \pm 4 \text{ \AA}$ ) is very similar to  $D_{\text{lam}} = 839 \pm 13 \text{ \AA}$ ; however, for  $D_{\text{red}} = 8.8$ , the repeat distance is increased by 10% compared to the bulk. For samples with intermediate molar masses ( $\bar{M}_n = 54.5$  and  $69.9 \text{ kg/mol}$ ), the AFM images do not show well-defined textures, which may be related to different lamellar orientations in the film. The lamellar orientations resulting from a number of samples as observed at the film surface using AFM – parallel for  $13.9$ – $54.5 \text{ kg/mol}$  and perpendicular for  $148$  and  $183 \text{ kg/mol}$  – are compiled in an orientation diagram as a function of the block copolymer molar mass and  $D_{\text{red}}$  in Fig. 2a. The solvent evaporation rate was found not to influence the general surface texture: A few samples were slowly cast from solution, i.e., the solvent was left to evaporate over a few days. The resulting lamellar orientation at the film surface was found to be the same as in the spin-coated samples (open symbols in Fig. 2a, [17]).

The surface topographies of the low and high molar mass samples are consistent with the parallel and the perpendicular lamellar orientation; however, a unique determination of the lamellar orientation *within* the film is impossible from the AFM images of the film surfaces alone [63]. Hence, we have performed GISAXS measurements in order to elucidate the inner film structure.

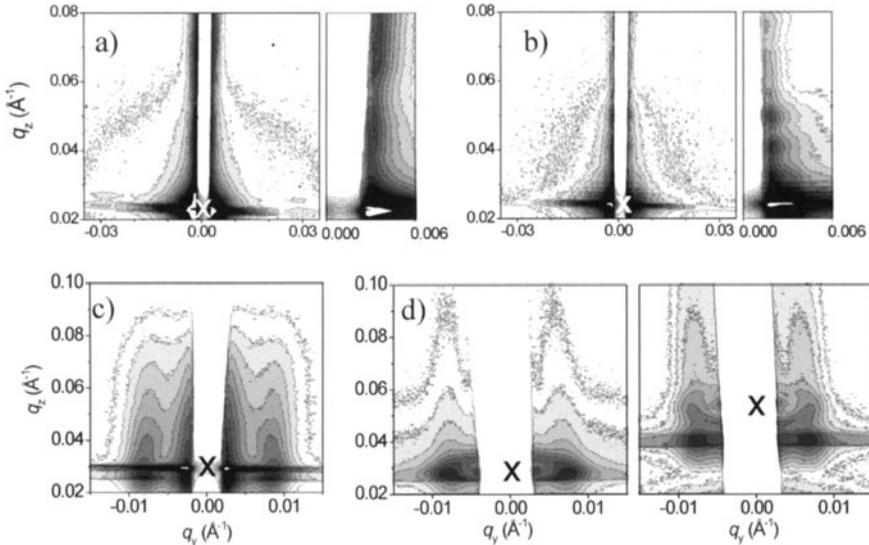
Scattering experiments were performed at beam line ID10B at the European Synchrotron Radiation Facility (ESRF) and at D-line at the Cornell High-Energy Synchrotron Source (CHESS). In both cases, a photon energy



**Fig. 2.** Lamellar orientation as a function of the molar mass and  $D_{\text{red}}$ . Results from AFM (a) and from GISAXS (b). The top-axis gives the  $D_{\text{lam}}$ -values from Ref. [64]. Triangles up: parallel, triangles down: perpendicular, diamonds: coexisting orientations, stars: weak surface texture, circles: no surface texture or scattering observed. (a) Filled symbols: spin-coated samples, open symbols: solvent-cast samples. (b) Filled symbols: non-treated films, open symbols: films were stained with OsO<sub>4</sub>

of 8 keV was used, i.e., the wavelength  $\lambda$  was 1.5 Å. Two collimation slits and a guard slit conditioned the incident beam. The vertical and horizontal width of the beam were 50 or 100 µm and 0.5 mm, respectively. The sample was mounted horizontally on a goniometer allowing adjustment of the incident angle of the X-ray beam. A 1.25 mm molybdenum rod mounted in front of the detector served as a beam stop for the intense reflected beam and the diffuse scattering. CCD cameras were used as two-dimensional detectors: At ID10B, the sample-detector distance was 0.71 m, and the pixel size of the lens-coupled CCD camera  $22.5 \times 22.5 \mu\text{m}^2$ , leading to a  $q$ -resolution of  $1.3 \times 10^{-4} \text{ \AA}^{-1}$  per pixel. At D-line, custom-built fiber-optically coupled CCD cameras [65] were used for the detection of the scattering intensity. The pixel size was 50.8 µm, which corresponds to  $1.7 \times 10^{-4} \text{ \AA}^{-1}$  per pixel. The dynamic range of both detectors extends from about 10 to  $5 \times 10^4$  counts per pixel and is well adjusted to the large range of scattering intensities. This set-up is able to resolve structures up to 1000 Å in size. The measuring times at both instruments were 5 or 10 sec.

The GISAXS intensity maps displays a wealth of information: Apart from the specularly reflected beam (partly hidden by the beam stop), the diffuse scattering containing information on the lamellar structure of the diblock copolymer films, is detected. The map of the thin, high molar mass film from Fig. 1c shows straight out-of-plane Bragg rods (Fig. 3c). Their  $q_y$ -positions are independent of  $q_z$  and correspond to  $2\pi/q_y = 762 \pm 46$  Å, a value 9% smaller than the bulk lamellar thickness of  $839 \pm 13$  Å. The intensity profile of the Bragg rods along  $q_z$  contains information on the extension of the lamellae



**Fig. 3.** GISAXS images from the same samples as in Fig. 1:  $\bar{M}_n = 22.6 \text{ kg/mol}$ ,  $D_{\text{lam}} = 197 \text{ \AA}$  (upper panel),  $\bar{M}_n = 183 \text{ kg/mol}$ ,  $D_{\text{lam}} = 839 \text{ \AA}$  (lower panel). Logarithmic intensity scales ranging from white to black. The white, vertical stripes at  $q_y \approx 0$  are due to the beam stop. The crosses indicate the position of the specularly reflected beam. (a)  $D_{\text{red}} = 2.1$ ,  $\alpha_i = 0.16^\circ$ , (b)  $D_{\text{red}} = 8.0$ ,  $\alpha_i = 0.17^\circ$ , (c)  $D_{\text{red}} = 2.8$ ,  $\alpha_i = 0.21^\circ$ , (d)  $D_{\text{red}} = 8.8$ ,  $\alpha_i = 0.20^\circ$  (left) and  $0.38^\circ$  (right). The narrow panels in (a) and (b) show regions of low  $q_y$ .

along the film normal, which, for ideally perpendicular lamellae throughout the film, is equal to the film thickness. For a quantitative analysis, one has to consider the dynamical effects cited above [59]. From the presence of the Bragg rods close to the expected positions, we conclude that the lamellae are perpendicular to the film surface throughout the film.

Straight Bragg rods at  $q_y$ -positions corresponding to a repeat distance of  $850 \pm 57 \text{ \AA}$  (a value 1 % larger than  $D_{\text{lam}}$ ) are present in the GISAXS intensity map of the thick, high molar mass film from Fig. 1d as well (Fig. 3d). In addition to these straight rods, a ring of high intensity is observed. In the map measured at  $\alpha_i = 0.38^\circ$  (right image in Fig. 3d), it is seen that the ring is centered around the specularly reflected beam.<sup>1</sup> Its radius corresponds to a repeat distance of  $\sim 900 \text{ \AA}$ , a value 8 % higher than  $D_{\text{lam}}$ , and we therefore attribute it to a Debye-Scherrer ring (well-known from transmission scattering from powders) around the specularly reflected beam. This indicates that lamellae with random orientation similar to the bulk state, are present in the interior of the film. This observation may explain the discrepancy of

<sup>1</sup> The direct beam is located at  $q_z = 0$ , a possible ring around the direct beam is shielded by the sample and can thus not be seen in the map.

the repeat distance at the film surface and the bulk lamellar thickness: Tilted lamellae near the film surface lead to an average repeat distance higher than  $D_{\text{lam}}$ .

The thick, low molar mass film from Fig. 1b (Fig. 3b) gives rise to completely different scattering: No Bragg rods are observed at the  $q_y$ -values corresponding to the bulk lamellar thickness ( $\pm 0.032 \text{ \AA}^{-1}$ ); instead, the intensity along  $q_z$  at  $q_z y \approx 0$  (close to the beam stop) shows peaks. Possible reasons are the roughness correlations between the film surface and the film/substrate interface [66,67] or diffuse Bragg sheets due to roughness correlations between lamellae with interfaces parallel to the film/substrate interface. We attribute the peaks to the presence of parallel lamellae, because for several GISAXS maps with  $\alpha_i = 0.18\text{--}0.21^\circ$ , the position of the peak at  $q_z = 0.05\text{--}0.06 \text{ \AA}^{-1}$  is close to the one predicted by the Born approximation for the first-order Bragg reflection of lamellae with  $D_{\text{lam}} = 197 \text{ \AA}$ . A detailed analysis requires calculations in the framework of the DWBA [59].

Similar scattering is observed from the thin, low molar mass film from Fig. 1a, also pointing to parallel lamellae (Fig. 3a). However, due to the low number of lamellae forming the film, the peaks at  $q_z y \approx 0$  are not as pronounced as for the thick films discussed above.

The lamellar orientations within the films as deduced from the GISAXS maps are compiled in Fig. 2b. For most samples, the same orientation as at the film surface (from AFM) is obtained using GISAXS: The lamellae are parallel for molar masses between 13.9–54.5 kg/mol and perpendicular for 148 and 183 kg/mol in a wide range of  $D_{\text{red}}$ -values. Only for thick, high molar mass films, the bulk limit seems to be reached, and other lamellar orientations are observed within the film.

## 4 Discussion

Thin, supported films of P(S-*b*-B) on Si/SiO<sub>x</sub> behave differently from the results reported in the literature on P(S-*b*-MMA) on Si/SiO<sub>x</sub>. In the last system, the lamellae have repeatedly been found to be parallel to the film surface for a large range of molar masses and film thicknesses [22,23,25,26,27,38,39,40,44]. The parallel orientation in P(S-*b*-MMA) can be explained by differences in interfacial tensions: The SiO<sub>x</sub>-layer is polar, and thus the interfacial tension of the polar PMMA block at the film/substrate interface is expected to be lower than the one of the rather non-polar PS block. In addition, the surface tension of PS is lower than the one of PMMA. Thus, the substrate is selective towards PMMA, whereas PS is preferentially located at the film surface. The fact that both film interfaces are selective to one of the blocks leads to the parallel orientation of the lamellae.

In further experiments on thin, supported, lamellar P(S-*b*-MMA) films, this hypothesis was tested by creating a non-selective substrate by coating Si/SiO<sub>x</sub> with a P(S-*r*-MMA) random copolymer [36,60]. Close to the sub-

strate, the lamellae were found to be perpendicular, whereas they were still parallel near the film surface. Additional coating of the surface of the P(S-*b*-MMA) film with P(S-*r*-MMA) leads to the perpendicular orientation also at the film surface [31,37]. These findings indicate that it is mainly the selectivity of the interfaces that determines the lamellar orientation up to a certain distance from the interface. However, entropic contributions to the free energy need to be considered in order to explain why the lamellae are perpendicular to non-selective surfaces and not randomly oriented [61,62].

The difference in surface tensions for the two blocks is similar for P(S-*b*-B) and P(S-*b*-MMA), thus the selectivity of the film surface cannot explain the different behavior of the lamellar orientation. However, in P(S-*b*-B), both blocks are rather non-polar, and thus, the Si/SiO<sub>x</sub> substrate can be considered to be close to non-selective. This may result in a change of the lamellar orientation in P(S-*b*-B) thin films from the perpendicular orientation for high molar mass films to the parallel orientation for low molar mass films. The free energy of a lamellar block copolymer melt near a hard wall given in [61] may explain the observed morphological transition [17]. Apart from the enthalpic contributions due to differences in surface or interfacial tensions of the two blocks (which can be considered to be independent of the block copolymer molar mass), entropic, molar-mass dependent terms need to be considered:

- The enrichment of chain ends near a hard wall lowers the free energy. The ends of the block copolymers are located mainly in the middle plane between two lamellar interfaces, and this effect thus stabilizes the parallel lamellar orientation with respect to the wall.
- The stretching of the copolymers normal to the lamellar interfaces (necessitated by the requirement of uniform density) is stabilized when the copolymers are stretched along the wall. This effect favors the perpendicular orientation.

A quantitative evaluation of the free energy for the specific case in question is, however, impossible because it contains coefficients which depend on the material parameters of the two blocks [61]. As many of these are unknown, we can only state that it is possible to find a set of coefficients which predicts the transition of lamellar orientation in P(S-*b*-B), in cases where the enthalpic term due to the selectivity of the wall to one of the blocks is negligible. If the latter condition is not fulfilled (such as in P(S-*b*-MMA) on Si/SiO<sub>x</sub>), it is not possible to find a set of parameters leading to the perpendicular orientation. Thus, the behavior of both P(S-*b*-B) and P(S-*b*-MMA) diblock copolymer films is qualitatively explained by this approach.

We conclude that in thin, supported films of lamellar P(S-*b*-B) on Si/SiO<sub>x</sub>, the lamellar orientation can be controlled by molar mass. This behavior can be explained by entropic contributions to the free energy of a lamellar block copolymer system near a hard wall and is in this framework consistent with the observations reported in the literature on thin films of P(S-*b*-MMA). GISAXS allowed us to complement AFM studies of the surface texture of

the films (and thus the lamellar orientation near the film surface) with information on the lateral and vertical structures within the film. Furthermore, it is interesting that in P(S-*b*-B), the perpendicular orientation is formed spontaneously – without any prepatterning of the substrate or outer fields, leading to surfaces structured on length scales of the order of 100 nm.

### Acknowledgements

We are indebted to F. Kremer, Universität Leipzig, for generous support and fruitful discussions. We thank B. Rheinländer, T. Bonné and A. Khaidarov, Universität Leipzig, for help with the measurements, E. Fontes, CHESS, for help with the setup at D-line, O. Konovalov for help at ID10B at ESRF. We appreciate stimulating discussions with I.I. Potemkin, Moscow State University, on polymer physics, and M. Rauscher, Max-Planck-Institut für Metallforschung, Stuttgart, on scattering theory. Financial support by Deutsche Forschungsgemeinschaft, Fonds der Chemischen Industrie and NATO is gratefully acknowledged. Part of this work is based upon research at the Cornell High-Energy Synchrotron Source (CHESS) which is supported by the National Science Foundation and the National Institute of Health/National Institute of General Medical Sciences under award DMR 0225180.

### References

1. I.W. Hamley, *Nanotechnology* **14**, R39 (2003).
2. M. Lazzari and M.A. López-Quintela, *Adv. Mater.* **15**, 1583 (2003).
3. M. Park, C. Harrison, P.M. Chaikin, R.A. Register and D.H. Adamson, *Science* **276**, 1401 (1997).
4. J.P. Spatz, P. Eibeck, S. Mößmer, M. Möller, T. Herzog and P. Ziemann, *Adv. Mater.* **10**, 849 (1998).
5. R. Glass, M. Möller and J.P. Spatz, *Nanotechnology* **14**, 1153 (2003).
6. T.L. Morkved, M. Lu, A.M. Urbas, E.E. Ehrichs, H.M. Jaeger, P. Mansky and T.P. Russell, *Science* **273**, 931 (1996).
7. T. Thurn-Albrecht, J. Schotter, G.A. Kästle, N. Emley, T. Shibauchi, L. Krusin-Elbaum, K. Guarini, C.T. Black, M.T. Tuominen and T.P. Russell, *Science* **290**, 2126 (2000).
8. L. Rockford, Y. Liu, P. Mansky, T.P. Russell, M. Yoon and S.G.J. Mochrie, *Phys. Rev. Lett.* **82**, 2602 (1999).
9. S.O. Kim, H.H. Solak, M.P. Stoykovich, N.J. Ferrier, J.J. de Pablo and P.F. Nealey, *Nature* **424**, 411 (2003).
10. T. Xu, J.T. Goldbach and T.P. Russell, *Macromolecules* **36**, 7296 (2003).
11. M. Kimura, M.J. Misner, T. Xu, S.H. Kim and T.P. Russell, *Langmuir* **19**, 9910 (2003).
12. A. Urbas, Y. Fink and E.L. Thomas, *Macromolecules* **32**, 4748 (1999).
13. V. Lauter-Pasyuk, H.J. Lauter, G.P. Gordeev, P. Müller-Buschbaum, B.P. Toporveg, M. Jernenkov, W. Petry, *Lamuir* **19**, 7783 (2003).
14. D.-M. Smilgies, P. Busch, C.M. Papadakis and D. Posselt, *Synchr. Rad. News* **15**, 35 (2002).

15. P. Busch, T. Bonné, D. Smilgies, D. Posselt and C.M. Papadakis, CHESS News Magazine 2002, p. 37.
16. P. Busch, D.-M. Smilgies, D. Posselt, F. Kremer and C.M. Papadakis, Macromol. Chem. Phys. **204**, F18 (2003).
17. P. Busch, D. Posselt, D.-M. Smilgies, B. Rheinländer, F. Kremer and C.M. Papadakis, Macromolecules **36**, 8717 (2003).
18. T.P. Russell, Mat. Sci. Rep. **5**, 171 (1990).
19. G. Krausch, Materials Science and Engineering **R14**, 1 (1995).
20. I.W. Hamley, *The Physics of Block Copolymers*, Oxford University Press 1998.
21. M. Tolan, *X-ray Scattering from Soft-Matter Thin Films. Materials Science and Basic Research*, Springer, Berlin 1999.
22. G. Coulon, D. Ausserre and T.P. Russell, J. Phys. France **51**, 777 (1990).
23. P. Bassereau, D. Brodbeck, T.P. Russell, H.R. Brown and K.R. Shull, Phys. Rev. Lett. **71**, 1716 (1993).
24. N. Koneripalli, F.S. Bates and G.H. Fredrickson, Phys. Rev. Lett. **81**, 1861 (1998).
25. L. Berger, V.A. Raghunathan, C. Launay, D. Ausserré and Y. Gallot, Eur. Phys. J. B **2**, 93 (1998).
26. A.P. Smith, J.F. Douglas, J.C. Meredith, E.J. Amis and A. Karim, Phys. Rev. Lett. **87**, 015503 (2001).
27. M. Maaloum, D. Ausserre, D. Chatenay, G. Coulon and Y. Gallot, Phys. Rev. Lett. **68**, 1575 (1992).
28. B.L. Carvalho and E.L. Thomas, Phys. Rev. Lett. **73**, 3321 (1994).
29. K.A. Orso and P.F. Green, Macromolecules **32**, 1087 (1999).
30. D.W. Schwark, D.L. Vezie, J.R. Reffner, E.L. Thomas and B.K. Annis, J. Mater. Sci. Lett. **11**, 352 (1992).
31. E. Huang, L. Rockford, T.P. Russell and C.J. Hawker, Nature **395**, 757 (1998).
32. D.L. Vezie, E.L. Thomas and W.W. Adams, Polymer **36**, 1761 (1995).
33. H. Hasegawa and T. Hashimoto, Macromolecules **18**, 589 (1985).
34. K. Ishizu and T. Fukuyama, Macromolecules **22**, 244 (1989).
35. A. Turturro, E. Gattiglia, P. Vacca and G.T. Viola, Polymer **36**, 3987 (1995).
36. P. Mansky, T.P. Russell, C.J. Hawker, M. Pitsikalis and J. Mays, Macromolecules **30**, 6810 (1997).
37. E. Huang, T.P. Russell, C. Harrison, P.M. Chaikin, R.A. Register, C.J. Hawker and J. Mays, Macromolecules **31**, 7641 (1998).
38. S.H. Anastasiadis, T.P. Russell, S.K. Satija and C.F. Majkrzak, Phys. Rev. Lett. **62**, 1852 (1989).
39. S.H. Anastasiadis, T.P. Russell, S.K. Satija and C.F. Majkrzak, J. Chem. Phys. **92**, 5677 (1990).
40. T.P. Russell, A. Menelle, S.H. Anastasiadis, S.K. Satija and C.F. Majkrzak, Macromolecules **24**, 6263 (1991).
41. M.D. Foster, M. Sikka, N. Singh, F.S. Bates, S.K. Satija and C.F. Majkrzak, J. Chem. Phys. **96**, 8605 (1992).
42. A. Menelle, T.P. Russell, S.H. Anastasiadis, S.K. Satija and C.F. Majkrzak, Phys. Rev. Lett. **68**, 67 (1992).
43. M. Sikka, N. Singh, A. Karim, F.S. Bates, S.K. Satija and C.F. Majkrzak, Phys. Rev. Lett. **70**, 307 (1993).
44. A.M. Mayes, T.P. Russell, P. Bassereau, S.M. Baker and G.S. Smith, Macromolecules **27**, 749 (1994).

45. P. Mansky, T.P. Russell, C.J. Hawker, J. Mays, D.C. Cook and S.K. Satija, Phys. Rev. Lett. **79**, 237 (1997).
46. J.R. Levine, J.B. Cohen, Y.W. Chung, P. Georgopoulos, J. Appl. Cryst. **22**, 528 (1989).
47. A. Naudon, in *Modern Aspects of Small-angle Scattering*, edited by H. Brumberger, (Kluver Academic Publishers, Amsterdam) 1995, p. 181.
48. P. Müller-Buschbaum, M. Wolkenhauer, O. Wunnicke, M. Stamm, R. Cubitt and W. Petry, Langmuir **17**, 5567 (2001).
49. P. Müller-Buschbaum, R. Cubitt and W. Petry, Langmuir **19**, 7778 (2003).
50. G. Vignaud, A. Gibaud, J. Wang, S.K. Sinha, J. Daillant, G. Grübel and Y. Gallot, J. Phys.: Condens. Matter **9**, L125 (1997).
51. P. Müller-Buschbaum, J.S. Gutmann, C. Lorenz-Haas, B. Mahltig, M. Stamm and W. Petry, Macromolecules **34**, 7463 (2001).
52. P. Müller-Buschbaum, J.S. Gutmann, C. Lorenz-Haas, O. Wunnicke, M. Stamm and W. Petry, Macromolecules **35**, 2017 (2002).
53. P. Müller-Buschbaum, Anal. Bioanal. Chem. **376**, 3 (2003).
54. A. Martorana, A. Longo, F. d'Acapito, C. Maurizio, E. Cattaruzza and F. Gonella, J. Appl. Cryst. **34**, 152 (2001).
55. S.K. Sinha, E.B. Sirota, S. Garoff and H.B. Stanley, Phys. Rev. B **38**, 2297 (1988).
56. V. Holý, J. Kuběna, I. Ohlidal, K. Lischka and W. Plotz, Phys. Rev. B **47**, 15896 (1993).
57. M. Rauscher, T. Salditt and H. Spohn, Phys. Rev. B **52**, 16855 (1995).
58. M. Rauscher, R. Panagio, H. Metzger, Z. Kovats, J. Domke, J. Peisl, H.-D. Pfannes, J. Schulze and I. Eisele, J. Appl. Phys. **86**, 6763 (1999).
59. P. Busch, M. Rauscher, D.-M. Smilgies, D. Posselt and C.M. Papadakis, in preparation.
60. E. Huang, P. Mansky, T.P. Russell, C. Harrison, P.M. Chaikin, R.A. Register, C.J. Hawker and J. Mays, Macromolecules **33**, 80 (2000).
61. G.T. Pickett, T.A. Witten and S.R. Nagel, Macromolecules **26**, 3194 (1993).
62. J.-U. Sommer, A. Hoffmann and A. Blumen, J. Chem. Phys. **111**, 3728 (1999).
63. M. Konrad, A. Knoll, G. Krausch and R. Magerle, Macromolecules **33**, 5518 (2000).
64. C.M. Papadakis, K. Almdal, K. Mortensen and D. Posselt, Europhys. Lett. **36**, 289 (1996).
65. M.W. Tate, E.F. Eikenberry, S.L. Barna, M.E. Wall, J.L. Lowrance and S.M. Gruner, J. Appl. Cryst. **28**, 196 (1995).
66. P. Müller-Buschbaum and M. Stamm, Macromolecules **31**, 3686 (1998).
67. J.S. Gutmann, P. Müller-Buschbaum, D.W. Schubert, N. Stribeck, D. Smilgies and M. Stamm, Physica B **283**, 40 (2000).

# III-V Semiconductors Under Ion Bombardment – Studied by RBS in situ at 15 K

E. Wendler and W. Wesch

Institut für Festkörperphysik, Friedrich-Schiller-Universität Jena  
Max-Wien-Platz 1, 07743 Jena, Germany

**Abstract.** The damaging behavior of various binary III-V compounds is studied in-situ at 15 K by Rutherford backscattering spectrometry, thus widely excluding the influence of thermal effects on damage formation. The fluence dependencies of the defect concentration measured show that the materials investigated can be classified in two groups with different behavior. In group A (InSb, InAs, InP, GaAs, GaP) a single ion impact produces heavily damaged and/or amorphous material, and amorphization of the implanted layers is reached by accumulation and stimulated growth of this material. In the case of group B (GaN, AlAs, AlN) only point defects and point defect complexes remain after a single ion impact. When the collision cascades start to overlap, these point defects recombine with those from previous ions, leading to an intermediate plateau of the defect concentration versus ion fluence at a rather low value. Finally, amorphization of these implanted layers is obtained by secondary processes (nucleation and growth of amorphous seeds) at very large fluencies.

## 1 Introduction

First papers on ion implantation into III-V semiconductors appeared in the early 1970ies, soon after the abilities of this technique were proven to create or to improve the conducting properties of silicon. In contrast to doping by diffusion, ion implantation has several advantages. Almost all ion species can be used and depth and concentration of the dopants introduced can be easily and separately controlled by the energy and the fluence of the implanted ions. Furthermore, much steeper gradients of the dopant concentration versus depth are possible with ion implantation than with diffusion and applying suitable energy- fluence implantation-sequences, a uniform dopant distribution can be obtained over large depths. And last but not least the use of implantation masks enables lateral structures with dimensions which are not accessible by diffusion processes.

However, ion implantation into semiconductors is inherently connected with the formation of radiation damage of the crystal lattice mainly due to the energy loss of the implanted ions into collision processes. Although there are few exceptions where defects themselves are involved to create special effects (for instance to form compensating centers to electrically isolate devices), in

most applications the ion implanted layers have to be annealed in order to obtain the desired electrical or optical properties.

In the past forty years much work has been carried out to study the damage evolution as a function of the implantation conditions (e.g. ion mass, ion energy, ion flux, target temperature) and to find out appropriate annealing conditions to reconstruct the crystal lattice and incorporate the foreign atoms on substitutional lattice sites. In the III-V compounds the defect annealing turned out to be much more complicated than in silicon, because more than one element occurs in the lattice cell. During the collision cascades local imbalances of the stoichiometry may occur, making perfect annealing more difficult. The consequence is that in III-V compounds the efficiency of certain annealing conditions depends much stronger on kind and concentration of defects produced during the implantation than in silicon.

Although the microscopic processes are not yet fully understood, for the most common III-V compounds (GaAs, InP, InAs and GaP) an empirical description of the ion-beam induced damage formation can be given in terms of critical temperatures  $T_c$  which follow from a defect-out-diffusion model firstly introduced by Morehead and Crowder [1] and later with a slightly different ansatz by W.J. Weber [2]. For a given ion/ion-energy/target combination the critical temperature is that temperature for which the ion fluence necessary for amorphization reaches infinity. If the implantation is performed at temperatures  $T_I$  well below  $T_c$ , amorphization is achieved by accumulation and overlapping of the heavily damaged and/or amorphous clusters produced by single ions. Contrary, at  $T_I \approx T_c$  the primary produced clusters dissociate, defects recombine and anneal, leaving behind a rather low defect concentration over a large range of ion fluences. In the case of the most extensively studied materials GaAs and InP, the  $T_c$  values correlate with well known annealing stages which can be attributed to the mobility of intrinsic defects (see e.g. [3] and references therein). This is in agreement with the defect-out-diffusion models, where the critical temperatures follow from. Further, this shows that the damage production during ion implantation at common temperatures between 77 K and about 600 K is strongly determined by thermal effects.

In contrast to the III-V materials mentioned above, the behavior of AlAs [4] and GaN [5] during ion implantation is very different, being dominated by huge recombination processes, the reason for which is not yet understood.

In order to study the ion-beam induced defect formation itself with almost no influence of thermal effects and to get more understanding about the differences between the various III-V compounds, ion irradiation and subsequent defect analysis have to be performed at low temperatures without changing the target temperature between implantation and measurement. Experiments of this type were already done on GaAs [6], AlAs [7] and GaN [8]. However, the conditions are too different to quantify and to compare the behavior of the materials. In this paper we present a comparative study of ion-beam induced

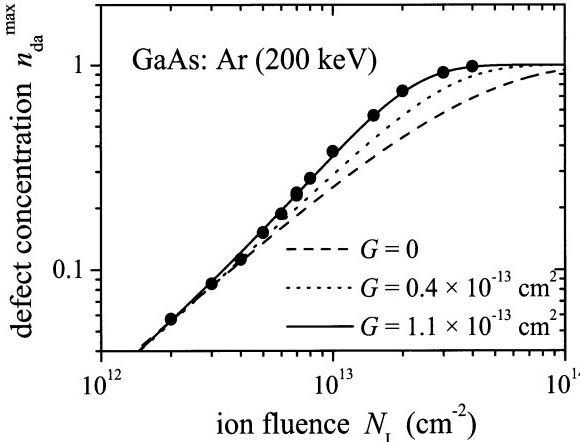
damage formation in various III-V compounds with both the implantation and the subsequent defect analysis by Rutherford backscattering spectrometry (RBS) in channeling configuration being performed at 15 K. The evolution of damage is analyzed applying the defect-interaction and amorphization model of Hecking et al. [9]. Special attention is put on the cross section of damage formation, which comprises information about the defect concentration remaining from the collision cascades of single ions. This cross section and consequently the sensitivity of the materials to ion-beam induced damage formation at 15 K is systematized using a force constant deduced from the long-wave optical properties [10].

## 2 Experimental Methods

In our investigations  $\langle 100 \rangle$  oriented InSb, InAs, InP, GaAs, GaP, AlAs and AlN was used. The available GaN was of  $\langle 0001 \rangle$  wurtzite type. Ion implantation and defect analysis by RBS were carried out in a special target chamber, which is connected to both the 400 kV implanter and the 3 MV Tandetron accelerator at the Institut für Festkörperphysik in Jena [11]. The sample holder is mounted on an automatic 4-axis goniometer. Cooling of the sample is realized with a He refrigerator, which is connected to the sample holder by copper cords, resulting in a target temperature of 15 K. The sample is subsequently moved in implantation position ( $7^\circ$  off axis) and after a certain ion fluence is implanted, it is tilted to the analyzing He beam for the RBS analysis. All experiments presented in this paper were performed at 15 K. The RBS analysis in channeling geometry was performed with 2 MeV He ions in the case of InAs and with 1.4 MeV He ions in all other cases. The angle of backscattering was  $170^\circ$ . The computer code DICADA [12] was applied to calculate the relative concentration of displaced lattice atoms  $n_{\text{da}}$  versus depth  $z$ . In this code the de-channeling background of the analyzing ions is treated analytically based on the discontinuous model of de-channeling [13]. In all cases a random distribution of the displacement distances of the displaced lattice atoms from their original sites was assumed. The thermal vibration of the atoms is treated with Debye's theory and for each material an appropriate Debye temperature was determined as described in [11]. In the following text the relative concentration of displaced lattice atoms  $n_{\text{da}}$  is for short referred to as defect concentration.

## 3 Mechanisms of Damage Formation

Our investigations of damage formation during ion implantation of various III-V compounds at 15 K showed that the materials investigated can be separated into two groups. In general similar results were found for InSb, InAs, InP, GaAs and GaP, which belong to a first group called A. A second group B is formed by GaN, AlAs and AlN. To exemplify the type A behavior, Sect. 3.1



**Fig. 1.** Relative concentration of displaced lattice atoms  $n_{da}^{max}$  occurring in the maximum of the distribution for 200 keV Ar ion implantation in GaAs. The lines are calculated using Eq. (1) and  $P = (2.9 \dots 2.5) \times 10^{-14} \text{ cm}^{-2}$

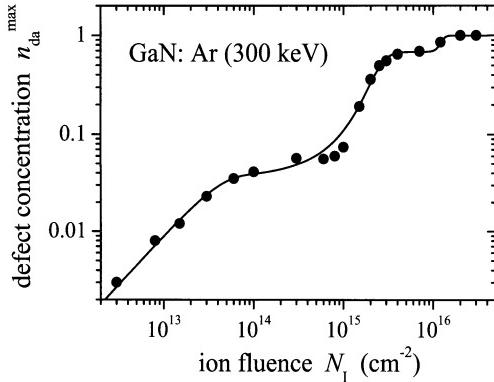
presents results for 200 keV Ar implantation in GaAs. The type B behavior is demonstrated in Sect. 3.2 for 300 keV Ar in GaN. The effect of the ion species on the defect concentration remaining after implantation is discussed for these two examples in Sect. 3.3.

### 3.1 Type A Materials – Example GaAs

From the RBS spectra measured in channeling direction after implantation of 200 keV Ar ions in GaAs at 15 K the defect distributions  $n_{da}(z)$  were calculated with the help of the DICADA code (not shown here, see [11]). The shape of these profiles does not change with increasing ion fluence  $N_I$  and it is in good agreement with that of the defect distribution calculated with the TRIM code (version TRIM97 was used [14]), see [11]. Figure 1 shows the defect concentration  $n_{da}^{max}$  occurring in the maximum of the distribution as a function of the ion fluence  $N_I$ . It can be seen that the defect concentration increases continuously until at  $N_I \approx 4 \times 10^{13} \text{ cm}^{-2}$  amorphization, i.e.  $n_{da}^{max} = 1$ , is reached. Similar results are found for Ar implantation into the other materials belonging to group A (for InAs see [15] and for GaP see [10]). A model to calculate the increase of the defect concentration with rising ion fluence is given by

$$\frac{dn_{da}^{max}}{dN_I} = (P + Gn_{da}^{max})(1 - n_{da}^{max}). \quad (1)$$

$P$  is the cross section for damage production per ion. The term  $Gn_{da}^{max}$  describes a stimulated growth of damage (cross section  $G$ ) occurring when already defects exist from previous ions. That means this growth takes place



**Fig. 2.** Relative concentration of displaced lattice atoms  $n_{\text{da}}^{\text{max}}$  occurring in the maximum of the distribution for 300 keV Ar ion implantation in GaN. The lines are calculated using Eq.(2) and the parameters given in Table 1

when the collision cascades from individual ions start to overlap. The dashed curve in Fig. 1 is calculated from Eq. (1) with  $P = 2.9 \times 10^{-14} \text{ cm}^2$  and  $G = 0$ , i.e. it is assumed that each ion produces a certain amount of damage. Larger values of  $P$  with  $G = 0$  shift the curve to lower fluences (not shown) and it is obvious from Fig. 1 that this simple assumption does not represent the experimental data. When also stimulated growth of damage is taken into account, i.e.  $G \neq 0$ , the experimental data can be well fitted with the help of Eq. (1) as can be seen from Fig. 1. Because the parameter  $P$  mainly results from the data at very low ion fluences and  $G$  has a marked influence on the calculated curves at higher fluences only (see Fig. 1), both parameters are well defined. The good agreement between the measured data and the calculated curve suggests that the two processes assumed (direct-impact production and stimulated growth of damage) are the main mechanisms yielding amorphization of the ion implanted GaAs layers. This holds well in all other type A materials.

### 3.2 Type B Materials – Example GaN

Amorphization of GaN occurs via three steps with two intermediate plateaus of the defect concentration, which is demonstrated in Fig. 2 plotting  $n_{\text{da}}^{\text{max}}(N_I)$  for 300 keV Ar implantation in GaN at 15 K [5]. Although implantation and defect analysis were performed at 15 K, amorphization is reached at very high ion fluences of  $N_I \approx 1 \times 10^{16} \text{ cm}^{-2}$  only. The complex structure of  $n_{\text{da}}^{\text{max}}(N_I)$  cannot be represented by Eq. (1), which assumes the existence of one type of damage (heavily damaged and/or amorphous clusters) only. However, also point defects and point defect complexes contribute to the yield of backscattered ions [16] and the measured defect concentration  $n_{\text{da}}$  can be taken as  $n_{\text{da}}^{\text{max}} = n_{\text{pd}} + n_{\text{a}}$  with  $n_{\text{pd}}$  and  $n_{\text{a}}$  being the (relative) concentration of point defects and amorphous regions, respectively. The evolution

of  $n_{\text{pd}}(N_I)$  and  $n_a(N_I)$  can be calculated using the defect-interaction and amorphization model of Hecking et al. [9] given by

$$\begin{aligned} \frac{dn_{\text{pd}}}{dN_I} &= P_{\text{pd}} e^{-R_{\text{pd}}^2 N_I^2} (1 - n_a) + C n_{\text{pd}}^{1.2} \left[ 1 - \frac{n_{\text{pd}}}{n_{\text{pd}}^c (1 - n_a)} \right] - \frac{dn_a}{dN_I} \frac{n_{\text{pd}}}{1 - n_a}, \\ \frac{dn_a}{dN_I} &= (P_a + G_a n_a) (1 - n_a). \end{aligned} \quad (2)$$

The model comprises the generation (cross section  $P_{\text{pd}}$ ) and recombination ( $R_{\text{pd}}$ ) of point defects, cluster formation ( $C$ ) to a saturation concentration  $n_{\text{pd}}^c$  of non-recombinable point defect clusters, generation ( $P_a$ ) and stimulated growth ( $G_a$ ) of amorphous material. It should be pointed out that  $P_{\text{pd}}$  is the cross section for defects remaining from the collision cascade of a single ion. Within the model (Eq. (2)) recombination means always recombination of defects resulting from different ions, thus taking place only when the collision cascades of the individual ions start to overlap. The curve in Fig. 2 is fitted to the experimental data with Eq. (2) using  $P_{\text{pd}}$  given in Table 1 (for all other data, which are not discussed in detail here, see [17]). Because of the clearly visible structure in the experimental data  $n_{\text{da}}^{\max}(N_I)$  the corresponding parameters are well defined. The good agreement between measured and calculated data indicates that the mechanisms behind Eq. (2) are well capable to explain the processes occurring during ion implantation of GaN. The linear increase of  $n_{\text{da}}^{\max}$  with  $N_I$  for  $N_I \leq 5 \times 10^{13} \text{ cm}^{-2}$  determines the cross section for point defect production  $P_{\text{pd}}$ . Only in that fluence range a good agreement between measured and calculated defect distributions is observed, whereas at higher fluences the profiles broaden towards the depth [5]. The first intermediate plateau of  $n_{\text{da}}^{\max}$  for  $1 \times 10^{14} \text{ cm}^{-2} \leq N_I \leq 1 \times 10^{15} \text{ cm}^{-2}$  results from the balance between production and recombination of point defects, thus yielding the cross section  $R_{\text{pd}}$ . The increase of  $n_{\text{da}}^{\max}$  for  $N_I \geq 1 \times 10^{15} \text{ cm}^{-2}$  is attributed to the formation of non-recombinable defect clusters which saturate at a concentration of  $n_{\text{da}}^c \approx 0.7$ . The analysis of the corresponding channeling spectra suggests the existence of both defect clusters and extended defects [17]. Finally amorphization starts at  $N_I \approx 1 \times 10^{16} \text{ cm}^{-2}$ . It is dominated by the stimulated growth of amorphous zones (because  $P_a \ll G_a$ , see [17]). The parameter  $P_a$  serves only to simulate the nucleation of amorphous seeds, which are produced by secondary processes and not within the direct ion impacts. A more detailed discussion of the mechanisms and its representation in Eq. (2) can be found in [17].

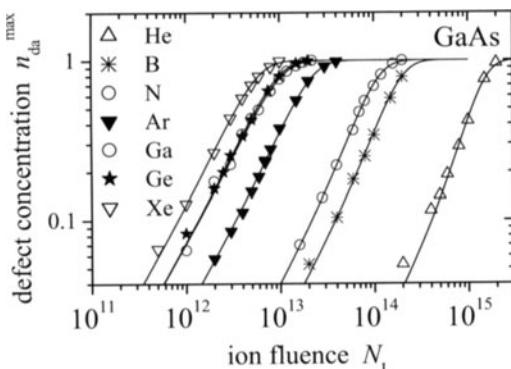
In case of AlAs amorphization by ion implantation at 15 K occurs in two steps only [18]. A second plateau as found in GaN does not exist. However, similar as in GaN a broad plateau at a very low level of  $n_{\text{da}}^{\max} \approx 0.17$  occurs and also amorphization takes place only at very high ion fluences of  $N_I \approx 1 \times 10^{16} \text{ cm}^{-2}$  within a similarly narrow fluence range as in GaN [4]. So besides the cluster formation, similar mechanisms are expected to occur in GaN and AlAs.

AlN was implanted at 15 K with 200 keV Ar ions up to a fluence of  $N_I = 2 \times 10^{15} \text{ cm}^{-2}$ . Within the investigated fluence range the defect concentration does not exceed a first plateau of  $n_{\text{da}}^{\text{max}} \approx 0.07$  [10]. AlN therefore also belongs to group B.

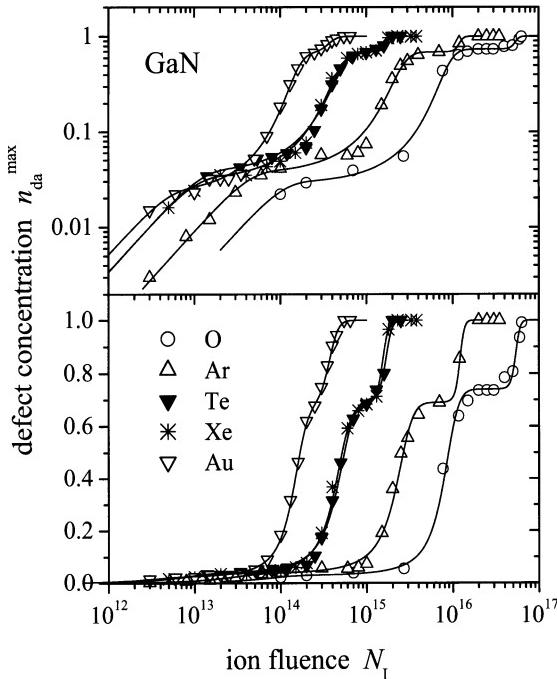
### 3.3 Effect of Ion Species

In order to demonstrate the effect of the ion species on the defect concentration remaining after implantation at 15 K, the defect concentration  $n_{\text{da}}^{\text{max}}$  measured in the maximum of the distribution is plotted versus the ion fluence  $N_I$  for different ion species implanted into GaAs (Fig. 3 and [3,19]) and GaN (Fig. 4 and [5,17]). Both figures clearly show that the general behavior (as demonstrated for Ar implantation in Sects 3.1 and 3.2) does not change, when different ion species are used. This suggests that the same mechanisms take place. The curves in Fig. 3 are fitted to the experimental data with Eq. (1) and those in Fig. 4 are fitted with Eq. (2).

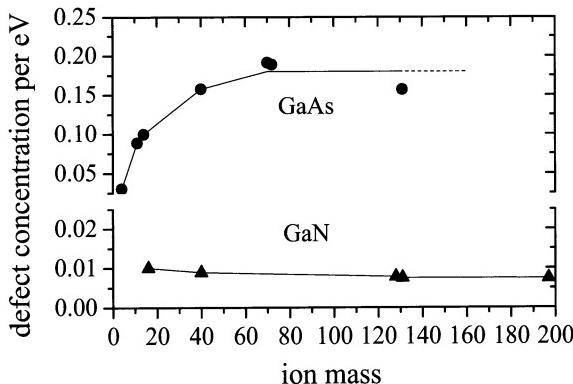
We now concentrate on the very early stage of damage formation, where the collision cascades from different ions do not overlap. The linear increase of  $n_{\text{da}}^{\text{max}}$  with  $N_I$  yields the cross section of defect production per ion,  $P$  (or  $P_{pd}$ , respectively). For a given substrate  $P$  has to increase with rising ion mass, because according to the collision kinematics more energy is deposited into collision processes. In order to clear  $P$  from this effect, it is recalculated into the defect concentration per energy deposited in collision processes. This concentration is given by  $P N_0 / (dE_{\text{recoil}} / dx)$  with  $N_0$  being the atomic density of the material and  $dE_{\text{recoil}} / dx$  the energy loss per ion and unit depth into collision processes (calculated with TRIM and taken in the maximum of the distribution). The resulting values are plotted in Fig. 5 for GaAs and GaN as a function of the ion mass.



**Fig. 3.** Relative concentration of displaced lattice atoms  $n_{\text{da}}^{\text{max}}$  occurring in the maximum of the distribution for different ion species implanted into GaAs. The lines are calculated using Eq. (1)



**Fig. 4.** Relative concentration of displaced lattice atoms  $n_{da}^{max}$  occurring in the maximum of the distribution for different ion species implanted into GaN. The lines are calculated using Eq. (2)



**Fig. 5.** Defect concentration per eV versus ion mass for ion implantation into GaAs (upper part) and GaN (lower part) at 15 K. Please notice the different scales of the ordinate for the two parts

In the case of GaN for all ion species an almost constant defect concentration per eV is obtained. This means that beside the collision kinematics

there is no influence of the ion mass on the defect concentration remaining from a single ion.

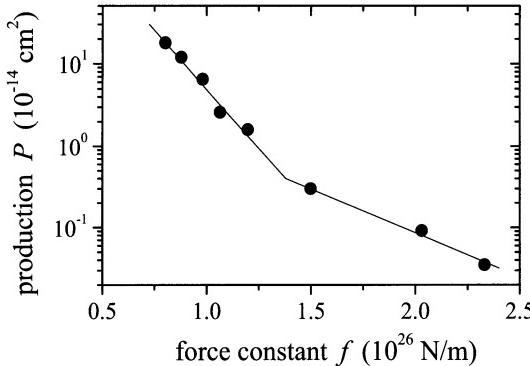
Contrary, in GaAs a constant defect concentration per eV is obtained only for ion masses above 40, whereas at lower masses the remaining defect concentration per eV strongly reduces (see Fig. 5). Consequently, there is an effect of the ion mass in addition to the purely kinematic one, which is connected with the density of the individual collision cascades, i.e. with the energy deposited per ion and unit depth [3]. For more dense collision cascades more damage remains than for more dilute ones.

#### 4 Systematization of Damage Production for all III-V Compounds Investigated

A comparison of the results for the various III-V compounds investigated is performed for the case of Ar ion implantation (see Table 1). From the fluence dependencies of the maximum defect concentration  $n_{da}^{\max}(N_I)$  the cross section of damage production per ion  $P$  (or  $P_{pd}$ , respectively) is obtained as shown in Sects 3.1 and 3.2. The resulting values are summarized in Table 1 and will be now regarded in more detail. It should be mentioned again that these values follow from the linear increase of  $n_{da}^{\max}$  with  $N_I$  at very low ion fluence, at which the collision cascades do not yet overlap, thus comprising information about the defect concentration remaining from the collision cascade of a single ion. Consequently, the variation of  $P$  for the different substrates should be related to some intrinsic material properties. Especially the strength of the bonds between the atoms is expected to be of main importance. In the case of the two-elemental materials under consideration there is an easy access to a force constant  $f$  via the long-wave optical parameters within the Reststrahlen-branch.  $f$  reads

$$f = \frac{m_1 m_2}{m_1 + m_2} \left[ \omega_T^2 + \frac{\varepsilon_\infty}{\varepsilon_\infty + 2} (\omega_L^2 - \omega_T^2) \right] \quad (3)$$

(for details of the derivation see [10]). Herein  $m_1$  and  $m_2$  are the atomic masses of the two elements of the crystal lattice,  $\omega_T$  and  $\omega_L$  are the transverse and longitudinal optical phonon frequencies and  $\varepsilon_\infty$  is the high-frequency dielectric constant of the corresponding material. The resulting values for  $f$  are given in Table 1. In Fig. 6 the cross section of defect production per ion  $P$  is plotted versus  $f$ . And indeed, a systematic dependence  $P(f)$  is found: for materials with low force constant the cross section is high and for those with high force constant it is low. From Fig. 6 it can be seen that the slope of  $P(f)$  changes at  $f_c \approx 1.4 \times 10^{26} N/m$  and it turns out that all materials belonging to group A exhibit values of  $f < f_c$  and those with  $f > f_c$  belong to group B (see Table 1). The main difference between the two groups was that, when the damage cascades start to overlap, in group A materials the stimulated growth of damage dominates, whereas in group B materials defect



**Fig. 6.** Cross section of damage production  $P$  versus the force constant  $f$  for Ar ion implantation into various III-V compounds at 15 K

recombination dominates (see Sects 3.1 and 3.2). To get an idea for this difference, it is necessary to regard  $P$  in more detail. It is given by

$$P = \int n_{da}^{ion}(A)dA \approx n_{da}^{ion} A_I. \quad (4)$$

$A$  is a plane parallel to the surface lying in the depth of maximum defect concentration (for only this is considered here, see Sect. 3.1) and  $n_{da}^{ion}(A)$  is the defect concentration produced by a single ion at that particular depth. Assuming this defect concentration to be constant over the area  $A$ , the most right side of Eq. (4) is obtained with  $n_{da}^{ion}$  being a mean value of the defect concentration over the area  $A_I$  damaged by one ion (projected to the surface). An estimate of  $A_I$  can be given in the following way. Stimulated growth of damage (or respective defect recombination) do only occur when the collision cascades start to overlap. Therefore, the ion fluence at the transition between linear accumulation and stimulated growth (or recombination)  $N_I^{trans}$  can be used to estimate  $A_I$  by  $1/N_I^{trans}$ . With (4)  $n_{da}^{ion}$  can then be estimated (see Table 1). In the case of type A materials  $0.5 < n_{da}^{ion} \leq 1$  is obtained. For such a high value heavily damaged and at least partially amorphized material must exist. Amorphous material can hardly crystallize at 15 K but grow due to the impact of subsequent ions. This explains why in the group A materials stimulated growth dominates. Contrary, in the case of type B materials the defect concentration produced per ion is rather low  $n_{da}^{ion} \approx 0.05 \dots 0.17$ . For such low concentrations only point defects are to be expected [16], which can annihilate when the cascades of the ions start to overlap. For instance the recombination of close Frenkel pairs seems to be possible even at such low temperatures [20]. So this explains why in these materials recombination of defects dominates.

Consequently, our results show that the marked different values of the defect concentration remaining from the collision cascades of single ions are the origin of the different behavior of the two groups of materials during

**Table 1.** Implantation conditions, force constant  $f$  in units  $10^{-26}$  N/m, cross section of damage formation  $P$ , transition fluence  $N_I^{\text{trans}}$  and defect concentration  $n_{\text{da}}^{\text{ion}}$  produced by a single ion. For detailed explanation of the quantities see text

substrate energy/ion	$f$	$P$ (cm $^2$ )	$N_I^{\text{trans}}$ (cm $^{-2}$ )	$n_{\text{da}}^{\text{ion}}$
AlN	200 keV Ar	$2.331 \times 10^{-16}$	$1.5 \times 10^{14}$	0.052
GaN	300 keV Ar	$2.031 \times 10^{-16}$	$6.6 \times 10^{13}$	0.062
AlAs	200 keV Ar	$1.499 \times 10^{-15}$	$5.7 \times 10^{13}$	0.171
GaP	200 keV Ar	$1.009 \times 10^{-14}$	$3.4 \times 10^{13}$	0.544
GaAs	200 keV Ar	$1.065 \times 10^{-14}$	$3.0 \times 10^{13}$	0.750
InP	350 keV Ca	$0.982 \times 10^{-14}$	$1.4 \times 10^{13}$	0.910
InAs	170 keV Ar	$0.924 \times 10^{-13}$	$9.0 \times 10^{12}$	1.080
InSb	300 keV Ar	$0.802 \times 10^{-13}$	$5.6 \times 10^{12}$	1.008

proceeding implantation. It remains open for further studies, why  $n_{\text{da}}^{\text{ion}}$  varies so strongly for the various materials. This question cannot be answered by MD simulations, because the available time scale is not large enough.

## 5 Summary

From our experimental data the defect concentration  $n_{\text{da}}^{\text{ion}}$  produced by a single ion impact is estimated.  $n_{\text{da}}^{\text{ion}}$  varies for the different materials and can be arranged by a force constant determined from the long-wave optical properties. The value of  $n_{\text{da}}^{\text{ion}}$  determines the behavior of the material during subsequent ion implantation. In the case of group A (InSb, InAs, InP, GaAs, GaP)  $n_{\text{da}}^{\text{ion}} > 0.5$  is obtained, suggesting amorphous material to remain after a single ion impact. Amorphization of the implanted layers is reached by accumulation and stimulated growth of the heavily damaged and at least partially amorphous regions produced by the single ions. In the case of group B (GaN, AlAs, AlN) the low value of  $n_{\text{da}}^{\text{ion}} < 0.2$  points to the existence of point defects and point defect complexes. These point defects recombine with those from previous ions, when the collision cascades start to overlap. This explains the occurrence of an intermediate plateau of the defect concentration versus ion fluence at a rather low value. Amorphization of the implanted layers is only obtained at very large fluences by secondary processes (nucleation and growth of amorphous seeds supported by the high concentration of implanted atoms).

## References

1. F.F. Morehead, B.L. Crowder, Radiat. Eff. **6**, 25 (1970).
2. W.J. Weber, L.M. Wang, Nucl. Instr. and Meth. B **91**, 63 (1994).
3. E. Wendler, B. Breeger, Ch. Schubert, W. Wesch, Nucl. Instr. and Meth. B **147**, 155 (1999).
4. E. Wendler, B. Breeger, W. Wesch, Nucl. Instr. and Meth. B **175-177**, 78 (2001).
5. E. Wendler, A. Kamarou, E. Alves, K. Gärtner, W. Wesch, Nucl. Instr. and Meth. B **206**, 1028
6. D.V. Stevanovic, N.P. Tognetti, G. Carter, Radiat. Eff. **71**, 95 (1983).
7. P. Partyka, R.S. Averback, D.V. Forbes J.J. Coleman, P. Ehrhart, W. Jäger, Appl. Phys. Lett. **65**, 421 (1994).
8. W. Jiang, W.J. Weber, S. Thevuthasan, J. Appl. Phys. **87**, 7671 (2000).
9. N. Hecking, K.F. Heidemann, E. TeKaat, Nucl. Instr. and Meth. B **15**, 760 (1986).
10. E. Wendler, L. Wendler, to be published in Phys. Rev. B
11. B. Breeger, E. Wendler, W. Trippensee, Ch. Schubert, W. Wesch, Nucl. Instr. and Meth. B **174**, 199 (2001).
12. K. Gärtner, Nucl. Instr. and Meth. B **132**, 147 (1997).
13. K. Gärtner, K. Hehl, G. Schlotzhauer, Nucl. Instr. and Meth. **216**, 275 (1983); Nucl. Instr. and Meth. B **4**, 55 and 63 (1984); **12**, 205 (1985).
14. J.F. Ziegler, J.P. Biersack, U. Littmark, The Stopping and Ranges of Ions in Solids, Pergamon, New York (1985).
15. W. Wesch, E. Wendler, M. Ridgeway, to be presented at IBMM 2004.
16. E. Wendler, W. Wesch, G. Götz, phys. stat. sol. (a) **112**, 289 (1989).
17. E. Wendler, A. Kamarou, E. Alves, W. Wesch, submitted to Phys. Rev. B.
18. E. Wendler, AIP Conference Proceedings **680**, 670 Melville, New York, 2003
19. B. Breeger, Ph.D. Thesis, Friedrich-Schiller-Universität, Jena (2000).
20. P. Partyka, Y. Zhong, K. Nordlund, R.S. Averback, I.M. Robertson, P. Ehrhart, Phys. Rev. B **64**, 235207 (2001).

# Size Controlled Si Nanocrystals

Margit Zacharias

Max Planck Institute of Microstructure Physics  
Weinberg 2, 06120 Halle, Germany

**Abstract.** A technique for tailor making silicon nanocrystals on 4-inch wafers has been developed. Following a standard procedure in silicon technology, a thermally unstable silicon compound in the form of an ultra-thin layer is first deposited on a substrate. A subsequent thermal treatment leads to a phase separation in this layer, in which silicon clusters and nanocrystals form depending on the temperature; these clusters and crystals are embedded in a matrix of thermally stable silicon dioxide. The size of the nanocrystals is controlled via the thickness of the deposited layer. This process makes the cost-effective manufacture of high-density arrays of silicon clusters or nanocrystals possible. The present paper summarizes our work on size controlled Si nanocrystals. The energy transfer to dopants, i.e. erbium, is presented. The results of our work are used to develop the optimization strategies for optical applications.

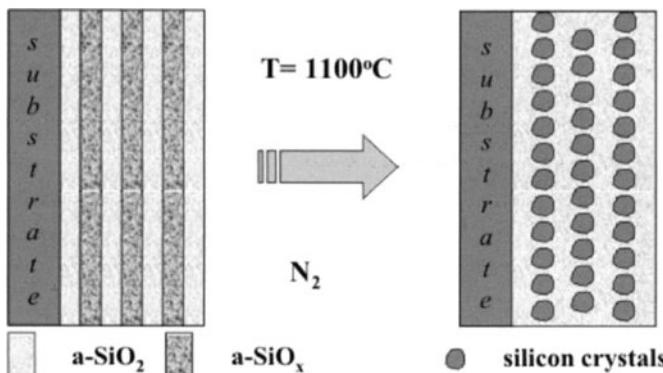
## 1 Introduction

Silicon, one of the base elements of our planet, is the foundation of the modern information society. Modern electronics would be unthinkable without the development of silicon transistors; such transistors are made possible only by the outstanding characteristics and stability of silicon and its oxides. However, the increasing miniaturization of microelectronics, the demands of optoelectronics, and the development of optical data transmission also show the limits of silicon technology: silicon is an indirect semiconductor and, as such, has a very inefficient light emission at room temperature. Thus, the structures most used in optoelectronics are based on the III-V elements such as gallium arsenide or indium phosphide, or corresponding combinations and are not compatible with silicon. Silicon is the standard material for modern microelectronics. Due to its indirect band structure it is not ideally suited for optoelectronic purposes. Nevertheless, because of its importance in microelectronics major research and development efforts have been undertaken to change silicon into a photonics material.

The discovery of room temperature photoluminescence from porous silicon [2,1] initiated an enormous re-interest for the luminescence of a silicon based material. Porous silicon was very extensively investigated and the preparation of thick films (100 µm and more) was realized which offered the possibility of a detailed optical characterization of the material. The external quantum efficiency of the material was reported to be up to 10%. Size

control can be realized by skillful choice of the etching parameter. The size distribution remains rather broad with randomly distributed nanowires and nanodots within the material. Porous silicon is not stable in luminescence due to the high reactivity of the large internal surfaces and hardly survives standard silicon technology procedures for device preparation. Hence, research focus shifted to better controlled growth techniques for the preparation of low dimensional Si based systems during the last years. The control of the size and density of Si quantum dots is of high interest for future applications. Low-dimensional silicon can be prepared as two-dimensional (2D) quantum wells, one-dimensional (1D) quantum wires and zero-dimensional (0D) quantum dots. Silicon quantum dots are more often referred to as silicon nanocrystals (nc-Si) due to the random crystal orientation normally obtained in growth. Such randomly distributed Si nanocrystals in amorphous  $\text{SiO}_2$  can be prepared for instance by Si implantation into thermal grown amorphous  $\text{SiO}_2$  [3]. A homogeneous distribution of the implanted Si ions can be realized if the  $\text{Si}^+$  ions are implanted with varied energies. Randomly distributed nanocrystals are formed by a subsequent annealing using temperatures in the range of 1000–1100°C. For lower temperatures amorphous clusters are formed. Size control was demonstrated by adjusting the implantation dose and annealing conditions. However, the half width of the size distribution is normally broad due to the Ostwald ripening process. The deposition of sub-stoichiometric  $\text{SiO}_x$  films represents a very simple method where the deposition can be done by sputtering [4], thermal evaporation [5] or PECVD [6]. Also in this case thermal annealing is used for phase separation of the  $\text{SiO}_x$  into the two stable components, Si and  $\text{SiO}_2$ . Size control was shown by adjusting the stoichiometry of the sub oxide. Both of the techniques are quite simple but the size distribution is dissatisfying especially for detailed optical measurements and for an industrial application.

In the following paper, a method is presented which allows for the first time a size controlled preparation of Si nanocrystals on whole Si wafers, an ordered layer arrangement, and a very high density of the crystals. The method allows a size controlled optical characterization of the nanocrystals as was demonstrated in [7] and represents a suitable solution for industrialized fabrication and application of such nanostructures. Erbium doped Si nanostructures are one of the most promising candidate for an on-chip fabrication of a silicon based light source. In addition, the wavelength of  $1.54\ \mu\text{m}$  of the Erbium ions fits well to the transmission maximum of the optical fibers used in telecommunication. Hence, the special properties of our Er doped Si nanocrystals and the strategies for future optimization are of high interest, too.



**Fig. 1.** Scheme of the preparation of size controlled Si nanocrystals using our new method

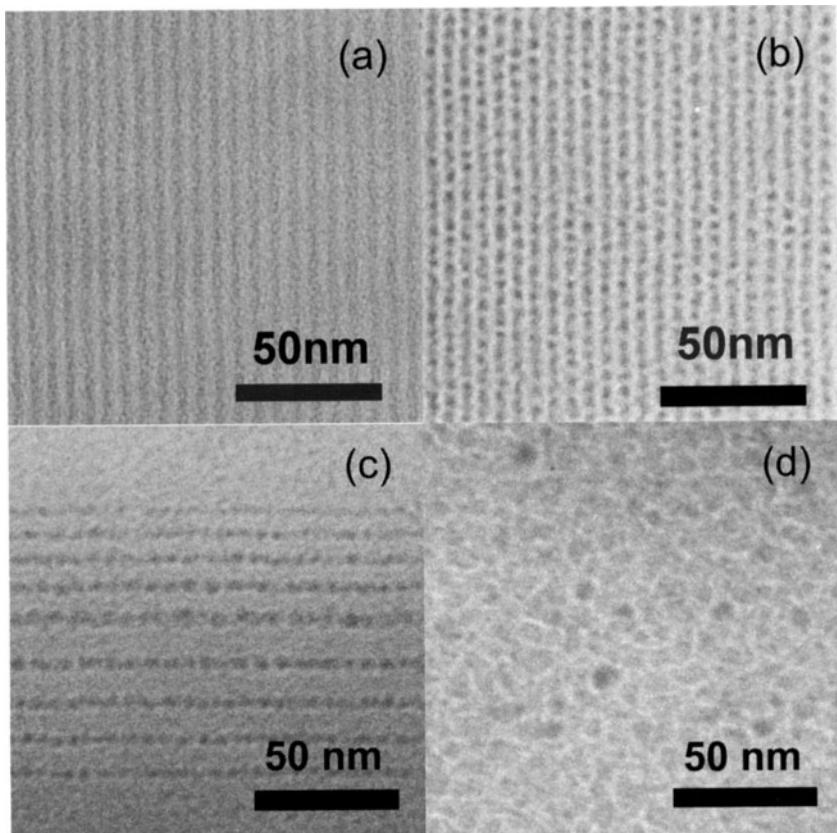
## 2 Experimental

### 2.1 Sample Preparation

The samples were prepared by reactive evaporation of SiO powder using a conventional thermal evaporation system which enables the use of 4 inch wafers. The material used for evaporation was a conventional SiO powder which was either evaporated under vacuum or by using additional oxygen gas ( $1 \times 10^{-4}$  mbar) for SiO<sub>2</sub> deposition. A shutter enabled the fast interruption of the process. The superlattice (SL) structure was prepared by alternatively depositing SiO or SiO<sub>2</sub> layers. The thickness of the SiO layers varied between 2 to 6 nm (depending on the sample). For the SiO<sub>2</sub> barrier layers a thickness of 4 nm was used. Usually, 30 to 45 periods were prepared. Sample rotation enabled a very high homogeneity over the whole wafer [8]. For crystallization the samples were annealed for 1 h at 1100°C under N<sub>2</sub> atmosphere resulting in a phase separation of the SiO layer into Si and SiO<sub>2</sub> and the crystallization of the amorphous Si cluster [9]. The size of the Si nanocrystals is determined by the SiO layer thickness, more details including the demonstration of using the CVD process are summarized in [10]. The method is schematically demonstrated in Fig. 1 (a). After crystallization selected samples were implanted with Erbium using an implantation energy of 300 keV. After implantation, the samples were annealed at 900°C under argon atmosphere. For comparison, a thick SiO<sub>2</sub> layer was implanted and annealed under the same conditions.

### 2.2 Measurements

Selected samples were characterized by transmission electron microscopy (TEM) using a CM20T or a JEM-4010 electron microscope. The cross section



**Fig. 2.** TEM image of (a) the as-prepared film, (b) the same film after annealing, (c) a film containing layers of different nc-sizes, and (d) a thick crystallized SiO film

samples were prepared in the usual way including final Ar ion milling. Imaging the superlattice structures as deposited and after annealing was realized applying the Fresnel defocus method at medium magnifications. A HeCd laser with power density below  $0.5 \text{ mW/cm}^2$  was used for the cw PL-measurements. The luminescence was detected either by a CCD camera for the visible range, or by a Germanium detector for the infrared luminescence. All spectra were corrected for spectral response of the measurement system.

### 3 Results

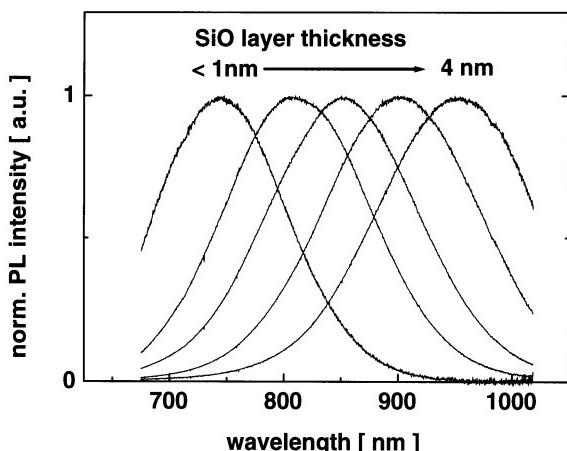
#### 3.1 Size, Density and Distribution Control

Figure 2 demonstrates samples realized by our special superlattice structure in comparison to a bulk SiO film in cross section TEM images. Fig. 2 (a)

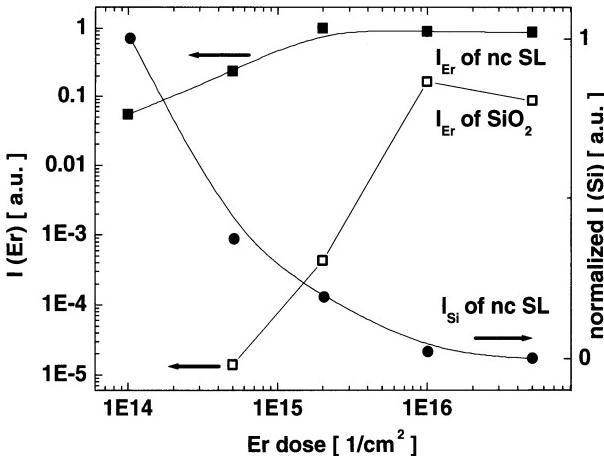
shows an amorphous as-prepared SL structure. After crystallization the ordered arranged nanocrystals are seen in the former SiO layers (Fig. 2 (b)). In addition, Fig. 2 (c) demonstrates a film prepared with varied SiO layer thicknesses, which results in a specific nanocrystal size in the respective depth of the film. For comparison, Fig. 2(d) represents a crystallized thick SiO film. As can be seen here, the film contains a broad distribution of nanocrystal sizes which are much larger than in the superlattice structure. The SL films were analyzed for photoluminescence (PL). The PL peak position shifts from around 930 nm to 730 nm depending on the size of the nanocrystals as demonstrated in Fig. 3 (a) on selected samples. The PL intensity increases linearly with the increase of the excitation power density. However, much more power density is needed if using a laser in the visible, i.e. an Argon laser, because of the lowering of the absorption cross section in the visible wavelength range by 5 orders of magnitude.

Figure 4 compares the luminescence intensity (nc-Si PL, Er PL) of sample pieces (having the same nanocrystal sizes by using one SL wafer) implanted with an increased Erbium dose. For comparison, a thick thermal grown  $\text{SiO}_2$  film implanted with the same doses is included into the figure. With increasing Er doping the luminescence of the Si nanocrystals is drastically reduced. The room temperature  $\text{Er}^{3+}$  PL at  $1.54 \mu\text{m}$  increases and saturates for an Er dose of around  $2 \times 10^{15} \text{ cm}^{-2}$ . Compared to the SL, much higher Er implantation is needed in case of the implanted bulk  $\text{SiO}_2$  without getting a similar PL intensity. A strong enhancement of the Er luminescence is found if using the Si nanocrystals as the mediating absorbing state. More details can be found in [11].

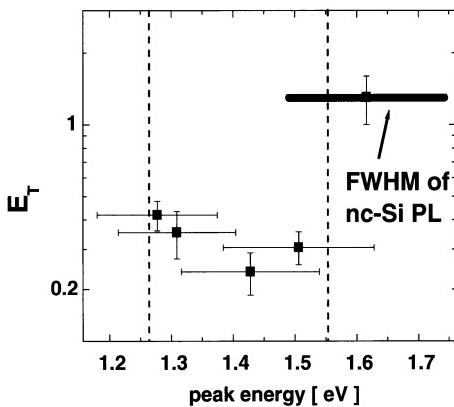
Figure 5 demonstrates the transfer efficiency for different sizes, i.e. confinement energies. The transfer efficiency can be estimated by measuring



**Fig. 3.** Photoluminescence of selected samples after crystallization demonstrating the size dependent blue shift



**Fig. 4.** PL intensity of the Si nanocrystals and the  $\text{Er}^{3+}$  state as a function of the Er dose. For comparison the behavior of a  $\text{SiO}_2$  film doped with the same dose is included



**Fig. 5.** Transfer efficiency to the  $\text{Er}^{3+}$  state as a function of the average size of the nanocrystals, i.e. PL peak energy

the respective PL intensities  $E_T \sim I(\text{Er})/(I(\text{nc-Si}) - I(\text{nc-Si, Er doped}))$ . The dashed horizontal lines represent the upper states of the  $\text{Er}^{3+}$  ions at  $0.8 \mu\text{m}$  and  $0.98 \mu\text{m}$ . The vertical bars represent the full width at half maxima of the nc-Si PL of the respective sample. As can be seen here, the transfer is enhanced only in case that the Si PL is strongly overlapping the discrete Er levels with an extension to the high energy side. The highest transfer is found for smallest sizes where the PL peak shows a very good overlapping with the levels at  $0.8 \mu\text{m}$ .

## 4 Discussion

### 4.1 Si Nanocrystals

The Si nanocrystals prepared with our new method of SiO/SiO<sub>2</sub> superlattices are nearly monodisperse. Size distributions with half width half maximum down to 0.2 nm were shown [10]. The method allows to synthesize very high densities of nanocrystals ( $\sim 10^{19} \text{ cm}^{-3}$ ) even for nanocrystal sizes below 2 nm. The missing defect luminescence, which is normally seen in Si implanted oxides around 750 nm, is a further hint for the outstanding quality of the prepared films. The phase separation was investigated in detail by combining IR and PL investigations [9]. Annealing at 900°C results in a complete phase separation into amorphous Si clusters embedded into the amorphous SiO<sub>2</sub> mask. These clusters are transferred to nanocrystals if using a temperature of 1100°C. The phase separation does not destroy the superlattice structure as can be seen in the TEM images (Fig. 2). All the nanocrystals are arranged in layers. Please note, that the images represent the overlay of nanocrystals from a depth of 10–20 nm.

Resonant photoluminescence measurements were used to prove the breakdown of the k-conservation rule for nc-Si by showing an increase in the no-phonon transition probability with decreasing crystal size. A no-phonon to phonon assisted transition probability ratio above 1 was detected at 4.5 K [7]. These results confirm quantum confinement as origin of the investigated luminescence signal. A separation of quantum confinement and migration effects on the PL signal was demonstrated which was possible due to the very narrow size distribution of the nc-Si and the detailed time and temperature dependent investigations of the photoluminescence. The PL peak position shift with the diameter (D) follows a  $D^{-0.6}$  dependence at room temperature which indicates that not only confinement effect but also additional processes are playing a main role. These effects are exciton migration and thermalization at room temperature as well as a possible better overlap of the electron and hole wave function in momentum space due to the spatial confinement into very small sizes following the Heisenberg uncertainty principle. In addition, an effect of a lowering of the barrier by a very thin SiO<sub>x</sub> layer around the nanocrystals might play a role, for more details please see [7].

### 4.2 Er Doping

Photoluminescence spectra of the undoped and Er-doped size-controlled nc-Si superlattices have been comparatively investigated as a function of implantation dose and Si nanocrystal size as demonstrated in Fig. 4. From investigations about the carrier dynamics we already estimated that one nanocrystal can pump 10–300 Er ion depending on the size of the nanocrystals [11].

The energy of the optically excited Si nanocrystals can almost be completely transferred to the Er<sup>3+</sup> ions, resulting in a very strong room temperature emission at 1.54 μm. The efficiency of the energy transfer increases for

smaller crystal sizes and an increasing overlap between the emission energy of the Si related luminescence and the absorbing  $\text{Er}^{3+}$  energy levels. For superlattices with nanocrystal sizes of 2 nm a transfer efficiency to the  $\text{Er}^{3+}$  ions is found to be nearly one at room temperature under an excitation with photons of 3.7 eV (Fig. 5). This property of a highly efficient excitation of the Er ions via the Si nanocrystals is in agreement with the Förster mechanism.

A Förster process can be understood as a dipole-dipole interaction between an excited "donor" and an "acceptor". The theory of this process was developed by Förster in 1948 [12]. Here the "donor" is represented by the excitons confined within the nanocrystals whereas the "acceptor" is represented by the  $\text{Er}^{3+}$  states. For achieving a Förster transfer between a donor and an acceptor a number of conditions should be fulfilled: (1) The donor and acceptor should have strong transitions in the ultraviolet to near infrared. (2) There should be a spectral overlap between the donor and acceptor states. (3) The distance between the donor and acceptor should be below the Förster zone. These conditions have to be met in optimizing the transfer process for applications.

One of the main advantages of our process is the possibility to prepare a very dense network of quasi monosized nanocrystals. These nanocrystals are highly luminescent and the luminescence intensity scales with the number of the nanocrystals. By using the developed size control we can shift the PL of the Si nanocrystals to get an optimum overlapping with Er levels by actually keeping the high density, i.e. luminescence, of the nanocrystals. As one can see in Fig. 5, the transfer efficiency is strongly depending on the spectral overlap of the nc-Si luminescence and the discrete levels of the  $\text{Er}^{3+}$  states. The best result is found for a sample containing  $\sim 2$  nm sized crystals with a PL peak position of 767 nm which is slightly blue shifted to the  $0.8\mu\text{m}$  state of the Er ions and hence has a very good overlap. The Förster zone was calculated for our system and represents a sphere with a radius of 10 nm around the donor, i.e. the Si nanocrystals. In our case, there are a number of nanocrystals within this sphere which contribute to the high efficiency of the process. In this way, the Er luminescence can be optimized for future applications in optical devices based on Er doped structures containing Si nanocrystals.

## 5 Conclusion

In conclusion, we demonstrated the preparation of quasi monosized Si-nc by using thermal evaporation and the preparation of a  $\text{SiO}/\text{SiO}_2$  superlattice structure on Si wafers. The light emitting Si nanocrystals were used to demonstrate a very efficient energy transfer from dense arrays of Si nanocrystals to  $\text{Er}^{3+}$  ions resulting in a room temperature luminescence at  $1.54\mu\text{m}$ . The conditions for optimizing future photonic applications based on the combination of Si nanocrystals and Er doping are discussed.

## Acknowledgements

The author gratefully acknowledge J. Heitmann for his very high effort during his PhD which contributed significantly to the work of my group, G. Lenk from Friedrich-Schiller-University Jena for performing Er ion implantation, and R. Scholz for TEM investigations. The work was financially supported by the Volkswagen Stiftung and the German Research Foundation (DFG).

## References

1. L. Canham, *Appl. Phys. Lett.* **57**, 1046 (1990).
2. V. Lehmann, and U. Gösele, *Appl. Phys. Lett.* **58**, 856 (1991).
3. T. Shimizu-Iwayama, N. Kurumado, D. E. Hole, and P. D. Townsend, *J. Appl. Phys.* **83**, 6018 (1998).
4. S. Hayashi, Y. Kanzawa, T. Kageyama, S. Takeoka, M. Fujii, and K. Yamamoto, *Sol. St. Commun.* **102**, 533 (1997).
5. U. Kahler and H. Hofmeister, *Opt. Mat.* **17**, 83 (2001).
6. F. Iacona, G. Franzò, and C. Spinella, *J. Appl. Phys.* **87**, 1295 (2000).
7. J. Heitmann, F. Müller, L.X. Li, M. Zacharias, D. Kovalev, and F. Eichhorn, *Phys. Rev. B*, **69**, 195309 (2004),
8. M. Zacharias, J. Heitmann, R. Scholz, U. Kahler, M. Schmidt, and J. Bläsing, *Appl. Phys. Lett.* **80**, 661 (2002).
9. L.X. Yi, J. Heitmann, R. Scholz, and M. Zacharias, *Appl. Phys. Lett.* **81**, 4248 (2002).
10. M. Zacharias, L.X. Yi, J. Heitmann, R. Scholz, M. Reiche, and U. Gösele, *Solid State Phenomena* **94**, 95 (2003).
11. J. Heitmann, M. Schmidt, L.X. Yi, M. Zacharias, V.Yu. Timoshenko, M.G. Lisachenko, and P.K. Kashkarov, *Materials science and engineering B* **105**, 214 (2003).
12. Th. Förster, *Ann. d. Physik* **6**, 55 (1948).

## Part VII

### **Disordered Systems and Soft Matter**

# Fat Tail Statistics and Beyond

Joachim Peinke<sup>1</sup>, Malte Siefert<sup>1</sup>, Stephan Barth<sup>1</sup>, Christoph Renner<sup>1</sup>,  
Falk Riess<sup>1</sup>, Matthias Wächter<sup>1</sup>, and Rudolf Friedrich<sup>2</sup>

<sup>1</sup> Institute of Physics, Carl von Ossietzky University of Oldenburg  
26111 Oldenburg, Germany

<sup>2</sup> Inst. Theor. Physik, University of Münster  
48149 Münster, Germany

**Abstract.** Based on data from three different systems, namely, turbulence, financial market and surface roughness we discuss methods to analyze their complexities. Scaling analysis and fat tail statistics in the context of Lévy distributions are compared with a stochastic method, for which a Fokker-Planck equation can be estimated from data. We show that the last method provides a more detailed characterization of complexity.

## 1 Introduction

Complex systems whose variables are governed by non Gaussian statistics, displaying fat tails, also called heavy tails, have gained considerable interest. On the one side, there is often the fundamental challenge to explain the mechanism leading to such a kind of statistics, as it is the case for turbulence, which is up to now one of the major unsolved problems of physics. On the other side, such statistics have very important consequences for applications, as it is the case for risk management.

In this contribution we want to focus mainly on the problem of fat tails in the statistics of data from turbulence, from financial market, and from surface roughness. In each case the question is to characterize the disorder of the system by the statistics of a scale resolving quantity,  $q(l, x)$ , where  $l$  denotes the scale and  $x$  the location. Usually, properties of the quantity  $q$  are analyzed for different fixed scales  $l$  over the whole space  $x$ . Typical points of interest are to what kind of statistics the probability density (distribution)  $p(q, l)$  belongs and what the functional  $l$  dependencies of general moments  $\langle q(l, x)^n \rangle$  are.

The question about the form of the distribution  $p(q, l)$  is directly linked to the topic of fat tail distributions. There is the famous approach by Lévy statistics. Variables governed by Lévy statistics belong to stable processes, i.e. the random variables and the sum over these random variables have the same form of statistics. The crucial point of Lévy statistics  $p_L(q)$  is that their fat tails do not decay exponentially fast as Gaussian distributions do. Lévy distributions decay with a power law for large or extreme events,

$$p_L(q) \cong \frac{1}{q^\beta} \quad \text{for large } q \quad \text{and} \quad 1 < \beta < 3. \quad (1)$$

It is easily seen that moments of the random variable  $\langle q^n \rangle$  diverge for  $n > \beta - 1$ . This is a quite unusual behavior. The reader may think over the implications of a system whose mean value is known but its variance is diverging. For example, what can we deduce from a results like the following: the probability of the magnitude of the next event is  $\langle q \rangle = \pm\infty$ . (For further reading we refer for example to [1].)

The other question of the functional  $l$  dependence of  $\langle q(l, x)^n \rangle$  is often connected with the question whether power law scalings is present. If  $\langle q(l, x)^n \rangle \propto l^{\xi_n}$  multiscaling or respectively multifractal properties are given. Per definition

$$\langle q^n \rangle = \int_{-\infty}^{\infty} dq q^n p(q) \quad (2)$$

the  $n$ th moment  $\langle q^n \rangle$  is directly linked to the probabilities  $p(q)$ . It is evident that if they obey Lévy laws, the question of multiscaling becomes inconsistent. For the case of turbulence we will show here, how these two aspects of multiscaling and fat tail distributions are connected.

So far we have reported on methods to analyze the disorder of systems for separated scales. In this paper we want to report on an approach, which enables to achieve a more complete characterization. Namely, we aim to achieve the knowledge of the joint probabilities of finding different values of  $q(l, x)$  for any scale  $l$ :  $p(q, l_1; q, l_2; \dots, q, l_n)$ . From this new approach we will see that the other approaches, like the multiscaling analysis, are not unique. For example, there are infinite many different ways to construct structures with the same multiscaling features.

## 2 Turbulence

The profound understanding of turbulence is up to now regarded as an unsolved problem. Although the basic equations of fluid dynamics, namely the Navier Stokes equations, are known for more than 150 years, a general solution of these equations for high Reynolds numbers, i.e. for turbulence, is not known. Even with the use of powerful computers no rigorous solutions can be obtained. Thus for a long time there has been the challenge to understand at least the complexity of an idealized turbulent situation, which is taken to be isotropic and homogeneous. This case will lead us to the well known intermittency problem of turbulence, which is nothing else than the occurrence of fat or heavy tailed statistics [5]. The central question is to understand the mechanism which leads to this anomalous statistics (see [2,3]).

The intermittency problem of turbulence can be reduced to the question about the statistics of the velocity differences over different distances  $l$ , measured by the increments  $q(l, x) = u(x + l) - u(x)$ . Usually the velocity increments are taken from the velocity component in direction of the distance vector  $l$ . These are the so-called longitudinal velocity increments. By the use

of energy considerations, a simple  $l$ -dependence of  $q(l, x)$  was proposed. It can be shown that the dissipation of energy takes place on small scales, namely, scales smaller than the so-called Taylor length  $\theta$ . On the other hand, the turbulence is generated by driving forces injecting energy into the flow on large scales,  $l \geq L_0$ , where  $L_0$  is given by the correlation length of  $u(x)$ . Thus for the transition from  $q(l, x)$  to  $q(l', x)$  with  $l' < l$ , the same amount of energy is transferred to a scale  $l$ , as it is transferred from this scale to even smaller scales  $l'$ . The conservation of the transferred energy (more precisely the energy rate per unit volume) can be assumed as long as  $L_0 > l, l' > \theta$ . This range is called the inertial range, where the turbulent field develops independently from boundary conditions and dissipation effects. It has been proposed that in this range universal features of turbulence arise.

Kolmogorov proposed that the disorder of turbulence expressed by the statistics of  $q(l, x)$  and its  $n$ th order moments  $\langle q(l, x)^n \rangle$  should depend only on transferred energy  $\epsilon$  and the scale  $l$ :  $\langle q(l, x)^n \rangle = f(\epsilon, l)$ . By simple dimensional arguments it follows that

$$\langle q(l, x)^n \rangle \propto \langle \epsilon^{n/3} \rangle l^{n/3}. \quad (3)$$

The simplest ansatz is to take  $\epsilon$  as a constant, thus the Kolmogorov scaling  $n/3$  is obtained [4]. Assuming a lognormal distribution for  $\epsilon$ , i.e. not  $\epsilon$  but  $\ln \epsilon$  has a Gaussian distribution, Kolmogorov and Oboukhov proposed for  $\langle \epsilon^{n/3} \rangle$  an additional scaling term, leading to the so-called intermittency [5] correction [6]

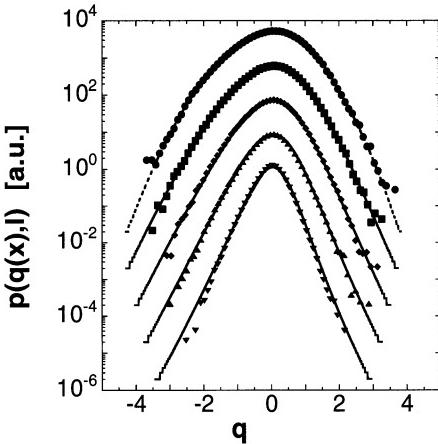
$$\langle q(l, x)^n \rangle \propto l^{\xi_n} \quad \text{with} \quad \xi_n = \frac{n}{3} - \mu \frac{n(n-3)}{18} \quad \text{and} \quad n > 2 \quad (4)$$

with  $0.25 < \mu < 0.5$  (for further details see [2]). The form of  $\xi_n$  has been heavily debated during the last decades.

Here we want to point out a general consequence of nonlinear scaling exponents  $\xi_n$ , namely, that the probability densities of  $p(q(l, x))$  cannot be Gaussian but must change their forms with  $l$ . This point is easily seen, if the definition of the moments Eq. (2) is considered. From the scaling relation Eq. (4), we take the scale dependence of the variance of the distributions:  $\sigma_l^2 := \langle q(l, x)^2 \rangle \propto l^{2\alpha}$  (for a general consideration,  $\alpha$  is any real number, for turbulence it is close to  $1/3$ ). Next, we express the integral of the  $n$ th moment (2) by the normalized variable  $\tilde{q}(l, x) = q/\sigma_l$

$$\langle q(l, x)^n \rangle = \sigma_l^n \int_{-\infty}^{\infty} \left( \frac{q}{\sigma_l} \right)^n p(q) dq = \sigma_l^n \int_{-\infty}^{\infty} (\tilde{q})^n \tilde{p}(\tilde{q}) d\tilde{q}. \quad (5)$$

If the normalized probabilities  $\tilde{p}(\tilde{q})$  do not change their form with  $l$ , like it would be the case if all probabilities are Gaussian, the integral on the right side is a constant and thus  $\langle q(l, x)^n \rangle \propto \sigma_l^n \propto l^{n\alpha}$  or the scaling indices  $\xi_n$  are linear in  $n$ . Saying this in another way, the nonlinear  $n$ -dependence of  $\xi_n$  means nothing else than that the probabilities  $p(q(l, x))$  must change their



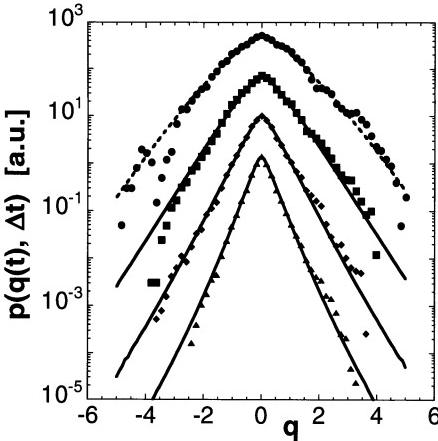
**Fig. 1.** Probability distributions obtained from a turbulent velocity signal measured in a free jet (bold symbols). The probabilities are compared with the numerical solution of the Fokker-Planck equation (solid lines). The scales  $l$  are (from top to bottom):  $l = L_0, 0.6L_0, 0.35L_0, 0.2L_0$  and  $0.1L_0$ . The distribution at the largest scale  $L_0$  was parameterized (dashed line) and used as initial condition for the Fokker-Planck equation ( $L_0$  is the correlation length of the turbulent velocity signal). The pdfs have been shifted in vertical direction for clarity of presentation and all pdfs have been normalized to their standard deviations; after [7]

form with the scale parameter  $l$ . In this way the discussion on the nonlinearity of  $\xi_n$  is linked to the discussion of non-Gaussian statistics. We see in Fig. 1 that for large scales ( $l \approx L_0$ ) the distributions are nearly Gaussian. As the scale decreases, the probability densities become more and more fat tailed.

Based on this finding, we can argue why a Lévy process, discussed above, is not consistent with the statistics of velocity increments of turbulence. As already mentioned, the Lévy statistics is based on the stability. Thus, the sum of the variables should have the same statistics. The sum of two successive increments  $q(l, x) + q(l, x + l) = (u(x + l) - u(x)) + (u(x + 2l) - u(x + l))$  is nothing else than the increment  $q(2l, x)$  of a scale twice as big. We have just seen that  $p(q(2l, x))$  has a different form, i.e. the process is not stable.

### 3 Finance

Next, we present some anomalous statistical features of data from the financial market. These features are astonishingly quite similar to the just discussed intermittency of turbulence [8] and are often called clustering of volatility (cf. [9,10,11]). The following analysis is based on a data set  $Y(t)$ , which consists of 1 472 241 quotes for US dollar-German Mark exchange rates from the years 1992 and 1993. Many of the features we will discuss here are also found in other financial data like for quotes of stocks. One central issue is the understanding of the statistics of price changes over a certain time interval  $\tau$  which determines losses and gains. The changes of a time series of quotations  $Y(t)$  are commonly measured by returns  $r(\tau, t) := Y(t + \tau)/Y(t)$ , logarithmic returns or by increments  $q(\tau, t) := Y(t + \tau) - Y(t)$ . The moments of these quantities display often power law behavior similar to the just



**Fig. 2.** Probability densities (*pdf*)  $p(q(t), \tau)$  of the price changes  $q(\tau, t) = Y(t + \tau) - Y(t)$  for the time delays  $\tau = 5120, 10240, 20480, 40960s$  (from bottom to top). Symbols: results obtained from the analysis of middle prices of bit-ask quotes for the US dollar-German Mark exchange rates from October 1st, 1992 until September 30th, 1993. Full lines: results from a numerical iteration of the Fokker-Planck equation (13); the probability distribution for  $\tau = 40960s$  (dashed lines) was taken as the initial condition. The pdfs have been shifted in vertical direction for clarity of presentation and all pdfs have been normalized to their standard deviations; after [12]

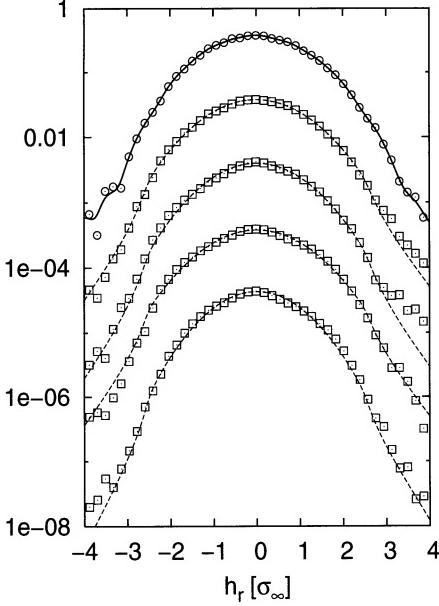
discussed Kolmogorov scaling for turbulence, cf. [8,13,14,15,16]. In addition one finds for the probability distributions an increasing tendency to fat tail probability distributions for small  $\tau$  (see Fig. 2) and [16]. This represents the high frequency dynamics of the financial market. The identification of the underlying process leading to these fat tail probability density functions of price changes is a prominent puzzle (see [9,14,15,17,18,19]), like it is for turbulence.

## 4 Surface Roughness

As a last example of structures whose complexity has attracted a considerable amount of interest we discuss the surface roughness. Among the techniques used to characterize scale dependent surface roughness, the most prominent ones are the concepts of self-affinity and multi-affinity, where the multi-affine  $f(\alpha)$  spectrum has been regarded as the most complete characterization of a surface [20,21,22,23]. Different definitions have been proposed to measure the scale dependent surface roughness. Here we are interested in the method of the analysis, thus we will proceed with a simple height increment,

$$q(l, x) := h(x + r/2) - h(x - r/2). \quad (6)$$

The meaning of centered, left or right justified, increments will be discussed in [24]. In Fig. 3 the scale dependent distributions of the height increments are shown, which display non Gaussian tails in a quite different structure of the fat tails (compare Figs. 1 and 2). The investigation based on multiscaling properties with respect to the moments of the height increments, suggest similarities to turbulence and financial data [24].



**Fig. 3.** Probability densities obtained from the surface roughness of a cobblestone road. Numerical solution of the integrated form of Fokker-Planck equation compared to empirical pdf (symbols) at different scales  $l$ . Solid line: empirical pdf parameterized at  $l = 188$  mm, dashed lines: reconstructed pdf. Scales  $l$  are 188, 158, 112, 79, 46 mm from top to bottom. Pdf are shifted in vertical direction for clarity of presentation and normalized to their standard deviation, after [27]

## 5 Stochastic Process for Scale Dependent Complexity

Based on the examples of complex structures and their analysis by means of the statistics of a scale dependent quantity  $q(l, x)$ , we want to report on an alternative, non conventional way to characterize anomalous statistics [7,12,25,26,27,28]. This method is based on the idea of a cascade process as discussed in the context of turbulence. In particular we show how from given data a stochastic process can be estimated. This makes it possible to model the measured statistics quite detailed. Guided by the finding that the statistics changes with scale, as shown in Figs. 1, 2 and 3, we consider the evolution of the quantity  $q(l, x)$ , or  $q(\tau, t)$  with the variable  $l$  or  $\tau$  respectively. Thus, we will consider processes which evolve in the scale variable. This we take as a stochastic cascade process.

For the following, we present a generalized discussion. Let us denote the given complex structure by  $Y(x)$ . With respect to the preceding sections,  $Y(x)$  denotes the velocity of a turbulent flow at the location  $x$ , the price of a stock at the time “ $x$ ”, or the height at  $x$ . To characterize the disorder of the structure  $Y(x)$  a local scale resolving quantity  $q(l, x)$  is considered. This quantity  $q$  is here defined as an increment  $q(l, x) := Y(x + l) - Y(x)$ , but may also be a wavelet, a local measure, or the sum of the square of derivatives of  $Y$ , just to mention some possibilities. For time series, like the financial market data,  $l$  stands for the time lag. Note, the moments of  $p(q)$  are called structure functions if  $q$  is an increment. It is easily seen that the second moment corresponds to the autocorrelation function  $\langle Y(x)Y(x+l) \rangle$ , which

is related by the Wiener – Khintchine theorem to the power spectrum of  $Y(x)$ . Higher order moments correspond to higher order two point correlations.

For one fixed scale  $l$  we get the scale dependent disorder by the statistics of  $q(l, x)$ . More complete stochastic information of the disorder on all length scales is given by the joint probability density function  $p(q_1, \dots, q_n)$ , where we set  $q_i = q(l_i, x)$ . Without loss of generality we take  $l_i < l_{i+1}$ . This joint probability may be seen in analogy to joint probabilities of statistical mechanics (thermodynamics), describing in the most general way the occupation probabilities of the microstates of  $n$  particles, where  $q$  is a six-dimensional state vector.

Next, the question is discussed whether it is possible to simplify the joint probability. In general, such joint probabilities can be expressed by conditional probabilities:

$$p(q_1, \dots, q_n) = p(q_1 | q_2, \dots, q_n)p(q_2 | q_3, \dots, q_n) \dots p(q_{n-1} | q_n)p(q_n). \quad (7)$$

A simplification is given, if for the multiple conditioned probabilities

$$p(q_i | q_{i+1}, \dots, q_n) = p(q_i | q_{i+1}) \quad (8)$$

holds. Eq. (8) is known as the condition for a Markov process evolving from state  $q_{i+1}$  to the state  $q_i$ , i.e. from scale  $l_{i+1}$  to  $l_i$ .

A “single particle approximation”, as used for the statistical mechanics of the ideal gas, would correspond to the condition:

$$p(q_1, \dots, q_n) = p(q_1)p(q_2) \dots p(q_n). \quad (9)$$

Based on equations (7) and (8), Eq. (9) holds if  $p(q_i | q_{i+1}) = p(q_i)$ . Only for the last case, the knowledge of  $p(q_i)$  is sufficient to characterize the complexity of the whole system. Otherwise, an analysis of only  $p(q_i)$ , for the scale  $l_i$ , is a non unique projection of the whole complexity of the system. Such a projection is done by the investigations of the moments as presented in Eqs. (3) and (4). This is also the case for the characterization of complex structures by means of fractality or multiaffinity (cf. for multiaffinity [22], for turbulence [2,3], for financial market [13]). A much more general characterization of the disorder is obtained if the transition probabilities from one scale to a smaller one are investigated and if equation (8) is fulfilled.

In the following, we focus on the more complete statistical characterization proposed above. The basic idea is to consider the evolution of  $q(l, x)$  at one fixed point  $x$  as a stochastic process in scale  $l$ . (The extensive discussion of the analysis of financial and turbulent data can be found in [7,12]).

To prove that a Markov process is given, it has to be shown that (8) holds. For a given data set  $Y(x)$  it is easy to evaluate the corresponding conditional probabilities by the calculation of  $q(l, x)$  at different points  $x$  for several fixed scales  $l_i$ . Having shown that the multiconditioned probabilities are equal to

the single conditioned probabilities, the evolution of the conditional probability density  $p(q, l | q_0, l_0)$  starting from the initial scale  $l_0$  follows

$$-l \frac{\partial}{\partial l} p(q, l | q_0, l_0) = \sum_{k=1}^{\infty} \left( -\frac{\partial}{\partial q} \right)^k D^{(k)}(q, l) p(q, l | q_0, l_0). \quad (10)$$

(The minus sign on the left side is introduced, because we consider processes running to smaller scales  $l$ , furthermore we multiply the stochastic equation by  $l$ , which leads to a new parameterization of the cascade by the variable  $\ln(1/l)$ , a simplification for a process with scaling law behavior of its moments.) This equation is known as the Kramers-Moyal expansion [29]. The Kramers-Moyal coefficients  $D^{(k)}(q, l)$  are defined as the limit  $\Delta l \rightarrow 0$  of the conditional moments  $M^{(k)}(q, l, \Delta l)$ :

$$\begin{aligned} D^{(k)}(q, l) &= \lim_{\Delta l \rightarrow 0} M^{(k)}(q, l, \Delta l), \\ M^{(k)}(q, l, \Delta l) &:= \frac{l}{k! \Delta l} \int_{-\infty}^{+\infty} (\tilde{q} - q)^k p(\tilde{q}, l - \Delta l | q, l) d\tilde{q}. \end{aligned} \quad (11)$$

Thus, for the estimation of the  $D^{(k)}$  coefficients it is only necessary to estimate the conditional probabilities  $p(\tilde{q}, l - \Delta l | q, l)$ . For a general stochastic process, all Kramers-Moyal coefficients are different from zero. According to Pawula's theorem, however, the Kramers-Moyal expansion stops after the second term, provided that the fourth order coefficient  $D^{(4)}(q, l)$  vanishes. In that case the Kramers-Moyal expansion is reduced to a Fokker-Planck equation (also known as Kolmogorov equation [30]):

$$\begin{aligned} -l \frac{\partial}{\partial l} p(q, l | q_0, l_0) &= \\ \left\{ -\frac{\partial}{\partial q} D^{(1)}(q, l) + \frac{\partial^2}{\partial q^2} D^{(2)}(q, l) \right\} p(q, l | q_0, l_0). \end{aligned} \quad (12)$$

$D^{(1)}$  is denoted as drift term,  $D^{(2)}$  as diffusion term. The probability density function  $p(q, l)$  obeys the same equation:

$$-l \frac{\partial}{\partial l} p(q, l) = \left\{ -\frac{\partial}{\partial q} D^{(1)}(q, l) + \frac{\partial^2}{\partial q^2} D^{(2)}(q, l) \right\} p(q, l). \quad (13)$$

We remind the reader that the Fokker-Planck equation describes the probability density function of a stochastic process generated by a corresponding Langevin equation (we use Itô's definition)

$$-\frac{\partial}{\partial l} q(l) = D^{(1)}(q, l)/l + \sqrt{D^{(2)}(q, l)/l} f(l) \quad (14)$$

where  $f(l)$  is white noise (Gaussian distributed and delta correlated). This feature provides an approach to simulate numerically typical data by the knowledge of the process coefficients, as it was proposed for example in [27] and shown in [28].

## 6 Discussion and Conclusion

Let us return to the discussion of scale dependent complexity in turbulence, finance and surface roughness. We have reported on a way to characterize this complexity by means of stochastic processes. A new point is that we consider processes running in the scale variable and not in time as usually. The presented features of the stochastic process have been well known for decades [30]. A strength of this analysis is that it does not depend on assumptions. By the estimation of the conditional probabilities, the Markovian property (8) as well as the Kramers-Moyal coefficients (11) can be evaluated [31,32]. Knowing the evolution equation (10), the n-increment statistics  $p(q_1, \dots, q_n)$  is known, too. Definitely, an information like the scaling behavior of the moments of  $q(l, x)$  can also be extracted from the knowledge of the process equations. Multiplying (10) by  $q^n$  and successively integrating over  $q_0$  and  $q$ , an equation for the moments is obtained:

$$-l \frac{\partial}{\partial l} \langle q^n \rangle = \sum_{k=1}^n \left( -\frac{\partial}{\partial q} \right)^k \frac{n!}{(n-k)!} \langle D^{(k)}(q, l) q^{n-k} \rangle. \quad (15)$$

Scaling, i.e. multi affinity, is obtained if  $D^{(k)}(q, l) \propto q^k$ , see [33].

Based on this procedure we were able to reconstruct directly from the given data the corresponding stochastic processes. Knowing these processes one can perform numerical solutions (see [7,12,27]). In Figs. 1, 2 and 3 the numerical solutions are shown by solid curves. We see that the heavy tailed structure of the probabilities is well described by this approach over a Fokker-Planck equation. This result may be taken as a verification of the above described method to reconstruct the process from pure data analysis.

The knowledge of the stochastic process equations allows a detailed analysis. For financial as well as for the turbulent data it was found that the diffusion term is quadratic in the scale resolved variable  $q$ . With respect to the corresponding Langevin equation (14), the multiplicative nature of the noise term becomes evident. It is this multiplicative noise which causes fat tails of the probability densities and eventually multifractal scaling. The scale dependency of drift and diffusion term corresponds to a non-stationary process in scale variables  $\tau$  and  $l$ , respectively. From this point we conclude and confirm our previous finding that a Lévy statistics for one fixed scale, i.e. for the statistics of  $q(x,l)$  for fixed  $l$ , is not the adequate class of statistics. The true structure of the complexity is unveiled by the two scale statistics.

Comparing the tips of the distributions at small scales in Figs. 1 and 2 a less sharp tip for turbulence is found. This finding is in accordance with a larger additive term in the diffusion term  $D^{(2)}$  for the turbulence data. Knowing that  $D^{(2)}$  has an additive and quadratic  $q$  dependence it is clear that for small  $q$  values, i.e. for the tips of the distribution, the additive term dominates. Taking this result in combination with the Langevin equation, we see that for small  $q$  values Gaussian noise is present, which leads to a Gaussian tip of the probability distribution, as found in Fig. 1. Thus we see that

the detailed analysis of the Fokker-Planck process allows to distinguish the complexity of turbulence and finance more profoundly, as it is the case for the multiscaling analysis where only different scaling indices are found. Here it should be noted, that the additive term in  $D^{(2)}$  violates the proposed proper scaling behavior for turbulence (4). In [34] it was even shown that the stochastic process will change considerable with the Reynolds number, putting into question concepts on universal turbulence.

As a last comment we want to point out, that the results for the surface roughness show, that also other forms of probability densities can be grasped quite well with this method. Thus the analysis with scale dependent stochastic process does not depend on scaling behavior or some special form of the probabilities. The basic features like Markov condition (8) can even be verified with the experimental data.

### Acknowledgements

We acknowledge helpful discussions with F. Böttcher St. Lück, and A. Naert.

### References

1. D. Sornette, Critical phenomena in natural sciences ((Springer, Berlin 2000)).
2. U. Frisch, Turbulence (Cambridge University Press, Cambridge, 1995).
3. K. R. Sreenivasan and R. A. Antonia, Annual Rev. Fluid Mech. **29**, 435(1997).
4. A. N. Kolmogorov, C.R. Acad. Sci. U.R.S.S., **30**, 301 (1941).
5. Here it should be noted that the term "intermittency" is used frequently in physics for different phenomena, and may cause confusions. This turbulent intermittency is not equal to the intermittency of chaos. There are also different intermittency phenomena introduced for turbulence. There is this intermittency due to the nonlinear scaling, there is the intermittency of switches between turbulent and laminar flows for non local isotropic fully developed turbulent flows, there is the intermittency due to the statistics of small scale turbulence which we discuss here as heavy tailed statistics.
6. A.N. Kolmogorov, J. Fluid Mech. **13**, 82 (1962).
7. Ch. Renner, J. Peinke and R. Friedrich, J. Fluid Mech **433**, 383(2000).
8. S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge, Nature **381**, 767 (1996).
9. P. Embrechts, C. Klüppelberg, and Th. Mikosch, Modelling extremal events (Springer, Berlin 2003).
10. J. Voit, The Statistical Mechanics of Financial Markets. Springer-Verlag, Heidelberg, 2001.
11. R. Mantegna & H.E. Stanley, An introduction to econophysics. Cambridge University Press, Cambridge, 1999.
12. Ch. Renner, J. Peinke and R. Friedrich, Physica **A 298**, 499 (2001).
13. J.-P. Bouchaud, M. Potters, and M. Meyer, European Phys. Journal B **13**, 595 (2000).
14. J.-P. Bouchaud, Quantitative Finance **1**, 105 (2001).

15. B. B. Mandelbrot, Quantitative Finance **1**, 113 (2001).
16. M. Karth, and J. Peinke, Complexity **8**, 34 (2002).
17. R. N. Mantegna, H. E. Stanley, Nature **376**, 46 (1995).
18. J. McCauley, Physica **A 285**, 506 (2000).
19. J.-F. Muzy, D. Sornette, J. Delour, and A. Arneodo, Quantitative Finance **1**, 131 (2001).
20. J. Feder, Fractals (Plenum Press, New York, London 1988).
21. F. Family and T. Vicsek (Editors), Dynamics of fractal surfaces, (World Scientific, Singapore 1991).
22. T. Vicsek, Fractal Growth Phenomena, (World Scientific, Singapore 1992).
23. A.-L. Barabási and H. E. Stanley( Editors), Fractal concepts in surface growth (Cambridge University Press, Cambridge 1995).
24. M. Waechter, aeXiv: physics/0404021.
25. R. Friedrich, J. Peinke & Ch. Renner, How to quantify deterministic and random influences on the statistics of the foreign exchange market. *Phys. Rev. Lett.*, **84** (2000), 5224–5227.
26. R. Friedrich & J. Peinke, *Physica D* **102** (1997), 147; R. Friedrich, & J. Peinke, *Phys. Rev. Lett.* **78** (1997), 863.
27. M. Wächter, F. Riess, H. Kantz, and J. Peinke, Europhys. Lett. **64**, 579 (2003).
28. G. R. Jafari, S. M. Fazeli, F. Ghasemi, S. M. Vaez Allaei, M. Reza Rahimi Tabar, A. Irajizad, and G. Kavei, Phys. Rev. Lett. **91**, 226101 (2003).
29. H. Risken, The Fokker-Planck Equation (Springer, Berlin, Heidelberg, 1989); N.G. Van Kampen, Stochastic processes in physics and chemistry (North Holland, Amsterdam, 1981); C.W. Gardiner, Handbook of Stochastic Methods, (Springer-Verlag Berlin, 1983); P. Hänggi & H. Thomas, Stochastic processes: Time evolution, symmetries and linear response *Physics Reports* **88** (1982), 207.
30. A. N. Kolmogorov, Mathematische Annalen **104**, 415 (1931).
31. M. Ragwitz and H. Kantz, Phys. Rev. Lett. **87??**, 254501(2001).
32. R. Friedrich and C. Renner and M. Siefert and J. Peinke, Phys. Rev. Lett.,**89**,149401 (2002).
33. R. Friedrich, Th. Galla, A. Naert, J. Peinke, and Th. Schimmel, Disordered Structures Analyzed by the Theory of Markov Processes, in J. Parisi, St. Müller, and W. Zimmermann, W. (eds.), Lecture Notes in Physics **503**, A Perspective Look at Nonlinear Media (Springer, Berlin, Heidelberg, 1998).
34. Ch. Renner, J. Peinke, R. Friedrich, O. Chanal, and B. Chabaud, Phys. Rev. Lett. **89**,124502 (2002).

# Statistical Physics for Complex Cosmic Structures

Luciano Pietronero<sup>1</sup> and Francesco Sylos Labini<sup>2</sup>

<sup>1</sup> Physics Department University of Rome “La Sapienza”  
P.le A. Moro 2, 00185 Rome Italy

<sup>2</sup> Université de Paris-Sud, Laboratoire de Physique Théorique Bâtiment 210  
F-91405 Orsay Cedex, France

**Abstract.** Cosmic structures at large scales represent the earliest and most extended form of matter condensation. In this lecture we review the application of the methods and concepts of modern statistical physics to these structures. This leads to a new perspective in the field which can be tested by the many new data which are appearing in the near future. In particular, galaxy structures show fractal correlation up to the present observational limits. The cosmic microwave background radiation, which should trace the initial conditions from which these structures have emerged through gravitational dynamics, is instead extremely smooth. Understanding the relation between the complex galaxy structures and the smooth microwave background radiation represents an extremely challenging problem in the field of structure formation.

## 1 Introduction

The specific issue which initiated our activity at the interface between statistical physics and cosmology, was the study of the clustering properties of galaxies as revealed by large redshift surveys, a context in which concepts of modern statistical physics (e.g. scale-invariance, fractality ...) found a ready application [1,2,3]. In recent years we have broadened considerably the range of problems in cosmology which we have addressed, treating in particular more theoretical issues about the statistical properties of standard primordial cosmological models [4,5,6] and the problems of formation of non-linear structures in gravitational N-body simulations [7,8]. What is common to all this activity is that it is informed by a perspective and methodology which is that of statistical physics: in fact we believe that this represents an exciting playground for statistical physics, and one which can bring new and useful insights into cosmology. In this lecture we briefly review the main points and we refer the interested reader to the recent monograph on the subject [4].

## 2 Galaxy Structures

The question which first stimulated our interest in cosmology has been the correlation properties of the observed distribution of galaxies and galaxy clusters [1,2,3]. It is here that the perspective of a statistical physicist, exposed to

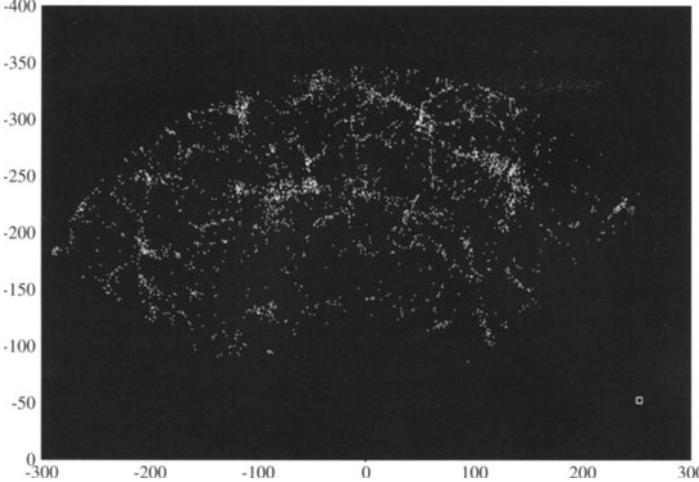
the developments of the last decades in the description of intrinsically irregular structures, is radically different from that of a cosmologist for whom the study of fluctuations means the study of small fluctuations *about a positive mean density*. And it is here therefore that the instruments used to describe strong irregularity, even if limited to a finite range of scales, offer a wider framework in which to approach the problem of how to characterize the correlations in galaxy distributions. Analyzing galaxy distributions in this wider context means treating these distributions without the *a priori* assumption of homogeneity, i.e. without the assumption that the finite sample considered gives, to a sufficiently good approximation, the true (non-zero) mean density of the underlying distribution of galaxies. While this is a simple and evident step for a statistical physicist, it can seem to be a radical one for a cosmologist [9,10]. After all the whole theoretical framework of cosmology (i.e. the Friedmann-Robertson-Walker – FRW – solutions of general relativity) is built on the assumption of an homogeneous and isotropic distribution of matter. Our approach is an empirical one, which surely is appropriate when faced with the characterization of data. Further it is evidently important for the formulation of theoretical explanations to understand and characterize the data correctly.

The maps probing the distribution of galaxies revealed in the first large three dimensional surveys, which were published in the eighties, structures – super-clusters, walls, voids, filaments – at scales much larger than had been suspected from the previous (fairly isotropic) angular data (see Fig. 1). The simple visual impression of such three-dimensional data - apparently showing large fluctuations up to the sample sizes – gives a strong *prime face* case for an analysis which does not assume the underlying distribution is uniform (at the scales probed), but rather encompasses the possibility that it may be intrinsically irregular. The quantitative results obtained with the standard analysis performed on these samples, which works with statistics which build in the assumption of homogeneity, give further evidence in this direction.

In simple terms, the standard statistical framework used in cosmology is the one developed for the studies of regular distributions as fluids: one assumes that the average density  $\langle n \rangle$  is well defined within a given sample and thus one focuses on the description of fluctuations around such a quantity (where  $\langle \dots \rangle$  is the unconditional ensemble average). In order to use a normalized (or reduced) two-point correlation function one defines

$$\xi(r) = \frac{\langle n(r)n(0) \rangle}{\langle n \rangle^2} - 1 \quad (1)$$

so that  $\xi(r) \neq 0$  implies the presence of correlations, while  $\xi(r) = 0$  is for a purely Poisson (uncorrelated) distribution: when  $\xi(r) \gg 1$  the regime is of strong-clustering while  $\xi(r) < 0$  describes anti-correlations. However such a definition does not include the possibility that the average density  $\langle n \rangle$  could



**Fig. 1.** This is a two dimensional slice of thickness 5 Mpc of a volume limited sample extracted from the first data release of the Sloan Digital Sky Survey [11]. The earth is in the point (0,0) and each point represent a galaxy. The empty region up to  $\sim 100$  Mpc is due to a selection effect and hence it is artificial. In this case there is no bias neither due to luminosity selection effects, nor to orthogonal projection distortions: Structures of galaxies of hundreds Mpc are still visible (distances are in Mpc – where we have taken the Hubble constant to be  $H_0 = 100$  km/sec/Mpc). The small square at the bottom right has a side of 5 Mpc, the typical clustering length according to the standard approach to the description of galaxy correlations [10] (from [4])

not be a well-defined quantity, i.e. that it could be a function of the sample size. This can be considered by using the conditional average density defined as

$$\langle n(r) \rangle_p = \frac{\langle n(r)n(0) \rangle}{\langle n \rangle} \quad (2)$$

which gives the average density around an occupied point of the distribution [1,2,?] (where  $\langle \dots \rangle_p$  is the conditional ensemble average). Clearly there is a simple relation which links these two functions

$$\langle n(r) \rangle_p = \langle n \rangle (1 + \xi(r)) ,$$

which is physically meaningful only in the case when the average density is a well defined quantity. Together with this simple change of statistical description of correlations one should consider however a subtle but important point related to the non-analytical character of fractal objects.

Fractal geometry, has allowed us to classify and study a large variety of structures in nature which are intrinsically irregular and self-similar [12]. The metric dimension is the most important concept introduced to describe these intrinsically irregular systems. Basically it measures the rate of increase of

the “mass” of the set with the size of the volume in which it is measured: Loosely speaking, it gives the total mass of an object as a function of its linear size. In terms of average density around an occupied point (averaged over all occupied points) if it scales as

$$\langle n(r) \rangle_p \sim r^{D-3}$$

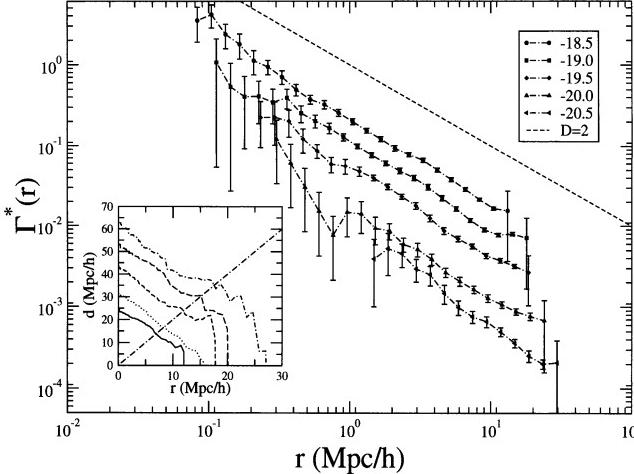
then the exponent  $D$  is called the *mass-length dimension* and it is  $0 < D \leq 3$  (in the three dimensional Euclidean space, where  $D = 3$  corresponds to the uniform case). In this situation the reduced correlation function (Eq.1) loses its physical meaning although its estimator can be defined in a finite sample. This is so, because its amplitude depends on the sample size as the average density. In fact one may show that in this case the estimator of  $\xi(r)$  in a finite sample of size  $R_s$  becomes (in an infinite volume it cannot be defined as a fractal) asymptotically empty, i.e.  $\langle n \rangle = 0$ )

$$\xi_E(r) = \frac{D}{3} \left( \frac{r}{R_s} \right)^{D-3} - 1 . \quad (3)$$

The use of the conditional average density we consider encompasses both some irregular distributions (simple fractals) and uniform distributions with small scale clusterization, in particular those described by standard cosmological models. If the sample size is much larger than the homogeneity scale  $\lambda_0$  one can detect using  $\langle n(r) \rangle_p$  the existence and location of  $\lambda_0$ . If, on the other hand the sample’s size is smaller than  $\lambda_0$  the conditional density allows the determination of the fractal exponent characterizing the clustering on these scales.

While the reduced correlation function  $\xi(r)$  has been found to show consistently a simple power-law behavior characterized by the same exponent in the regime of strong clustering, there is very considerable variation between samples, with different depth and luminosity cuts, in the measured *amplitude* of mean-density-normalized correlation function [13]. This variation in amplitude is usually ascribed *a posteriori* to an intrinsic difference in the correlation properties of galaxies of different luminosity (“luminosity bias”, or “luminosity segregation”). It may, however, have a much simpler explanation in the context of irregular distributions. As mentioned above, in a simple fractal (at least within the sample scale), for example, the density in a finite sample decreases on average as a function of sample size; samples of increasingly bright galaxies are in fact generically of greater mean depth, which corresponds to an increasing amplitude of the correlation function  $\xi(r)$  normalized to the “apparent” average density in each sample (see Eq.3).

With this motivation we have applied simple statistical methods which allow a characterization of galaxy clustering, irrespective of whether the underlying galaxy distribution is homogeneous or irregular at the sample size. Up to scales of tens of Mpc, we have found that galaxy distributions in many different surveys [2,3] show fractal properties corresponding to  $D \approx 2$  with



**Fig. 2.** The estimator of the conditional average density  $\Gamma^*(r)$  in the different samples of the CfA2-South galaxy catalog. (Adapted from [3])

no robust evidence for homogeneity (see Fig. 2). It is on this important issue that future surveys (e.g. the Sloan Digital Sky Survey – SDSS — [11]) will allow us to place much tighter constraints. We stress however the following important points: if the homogeneity scale is determined to be finite, the fractal-inspired analysis remains valid in two respects: (i) it is the unique way to detect the homogeneity scale itself without a priori assumptions, and (ii) it gives the right framework to obtain a full geometrical and statistical description of the strongly clustered region [4]. This result has caused a passionate debate in the field because it is in contrast with the usual assumption of large scale homogeneity which is at the basis of most theories. Our basic conclusion is that, given the fractal nature of galaxy clustering on small scales, one should change the general perspective with respect to the problem of non-linear structure formation. Note that this is the case even it turns out that there is crossover to homogeneity because, at least in a certain range of scales, that in which there is strong clustering, structures have a fractal nature and this cannot be described as perturbations to a smooth fluid. It can be interesting to note that some cosmologists (e.g. [9]) have confirmed our measurements about fractality at scales of order tens Mpc, the debate about the transition to homogeneity concerning scales of one order of magnitude larger. The discussion about the homogeneity scale is indeed an hot topic in contemporary science, touching many deep and important key-issues in different fields of physics, from large scale astrophysics to gravitation, to the question of the nature and amount of dark matter. To these topics it is dedicated a recent book [15]: Its main aim is in fact to address, in a novel and open way, the problem of structures in cosmology, with a reach historical per-

spective and stressing the open questions which arise from the consideration of their intrinsic complexity.

One of the reason of the resistance to our ideas in cosmology, is that we do not present neither an alternative theory or a model nor a physical theory for the fractal growth phenomena in the gravitational case: this is due to the fact that this issue is indeed extremely difficult as any out-of-equilibrium irreversible dynamical growth process. Actually the study of such processes is at the frontier of the theoretical efforts in statistical physics. Let us, however, note briefly two common *theoretical* objections to this approach. One is that fractals are incompatible with what is called the “cosmological principle”, by which it is meant that there is no privileged point or direction in the universe. This is a simple misconception about fractals. These irregular distributions are in keeping with this principle in exactly the way as inhomogeneities treated in the standard framework of perturbed FRW models, i.e. they are *statistically* stationary and isotropic distributions in space in the sense that their statistical properties are invariant under translation and rotation in space. Another common objection is that there is an inconsistency in using the Hubble law, which is used to convert redshift to physical distance, if one does not assume homogeneity. This objection forgets that the Hubble law is an established empirical relation, independent of theories explaining it. Further it can be noted that what we are concerned with is the distribution of *visible* matter: given that standard cosmological models describe a universe whose energy density is completely dominated by several non-visible components, reconciling the two is not in principle impossible [14].

### 3 Primordial Density Fields

As already mentioned, the standard interpretation of galaxy and cluster correlations results from the  $\xi(r)$  analysis and can be summarized as follows: (i) Power-law behavior in the regime of strong clustering ( $\xi \gg 1$ ) (ii) Amplitude which changes from sample to sample. These observations of the varying amplitude are ascribed to real physics, rather than being simply finite size effects as in the alternative explanation we present. Generically this goes under the name of “bias”, which is then simply taken to mean any difference between the correlation properties of any class of object (galaxy of different luminosity or morphology, galaxy cluster, quasar...) and that of the “cold” dark matter (CDM) which, in standard models, dominates the gravitational clustering dynamics in the universe. The case that such a difference manifests itself as an overall normalization of  $\xi(r)$ , as in the observations we have been discussing, is known as “linear bias”. In order to understand the problematic aspects of this interpretation, one should consider an important feature common to all standard theoretical models of cosmological density fields, like CDM.

First let us briefly the main properties of these fields. Matter distribution in the early universe is supposed to be well represented by a continuous Gaussian field with appropriate correlation properties. According to the standard picture of the Hot-Big-Bang scenario [10] matter density fluctuations had interact with radiation at early times leaving therefore an imprint of its correlation properties of the cosmic microwave background radiation (CMBR) anisotropies, in particular determining an important property at large scales. This is the so-called Harrison-Zeldovich (HZ) condition on the power spectrum (PS)  $P(k)$  (Fourier Transform of  $\xi(r)$ ) of these tiny fluctuations with respect to perfect uniformity [6]:  $P(k) \sim k$  at small  $k$  (large scales). We have noticed in [6] that this condition implies that their primary characteristic is to show surface mass fluctuations, which are the most depressed fluctuations possible for any stochastic distribution. In a simple classification of all stationary stochastic processes into three categories, we highlighted with the name “super-homogeneous” the properties of the class to which models like this belong: let us briefly discuss this point.

A simple description of these systems can be the following. An uncorrelated Poisson system (or with positive and finite range of correlations) has at large scale  $\xi(r) = 0$  and hence the PS is  $P(k) \sim \text{const}$ : for this reason it satisfies the condition

$$\int_0^\infty \xi(r)r^2 dr \sim P(0) = \text{const}.$$

In this situation mass fluctuations in spheres grows as  $\langle \Delta M^2(R) \rangle \sim R^3$  (Poisson fluctuations): this is, for example a perfect gas at thermodynamical equilibrium. On the other hand, one may find systems with a power-law correlation function as in thermodynamical critical phenomena. In this situation  $\xi(r) \sim r^{-\gamma}$  where  $0 < \gamma < 3$ ,  $P(k) \sim k^{-(3-\gamma)}$  and

$$\int_0^\infty \xi(r)r^2 dr \sim P(0) = \infty :$$

The correlation length diverges and the system has critical features that is mass fluctuations growth faster than in the Poisson case, i.e.  $\langle \Delta M^2(R) \rangle \sim R^{6-\gamma}$ . Finally there are systems such that

$$\int_0^\infty \xi(r)r^2 dr \sim P(0) = 0 \tag{4}$$

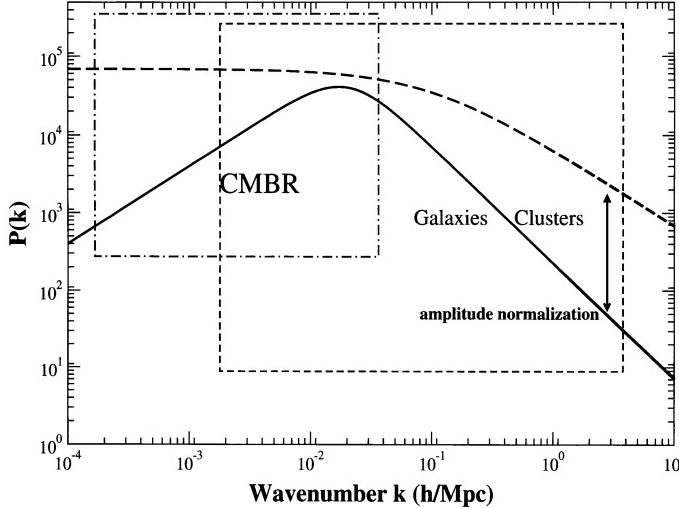
and mass fluctuations in spheres go as  $\langle \Delta M^2(R) \rangle \sim R^2$ : In statistical physics language they are well described as glass-like or long-range ordered configurations. For example the generation of points distributions in three dimensions with properties similar to HZ ones encountered in statistical physics is represented by the one-component system of charged particles in a uniform background [5]. This latter system, appropriately modified, can produce equilibrium correlations of the kind assumed in the cosmological context. These

systems are characterized by a long-range balance between correlations and anti-correlations (Eq.4) and show the most depressed possible fluctuations in any stochastic distribution [6].

The HZ type spectrum was first given a special importance in cosmology with arguments for its “naturalness” as an initial condition for fluctuations in the framework of the expanding universe cosmology. Basically it satisfies the global condition on large scale mass fluctuations  $P(0) = 0$ , which is the only one compatible with the FRW metric, as it avoids the divergence of the fluctuations of the gravitational potential (as it happens in the simple Poisson case [6]). It subsequently gained in importance with the advent of inflationary models in the eighties, and the demonstration that such models quite generically predict a spectrum of fluctuations of this type. Since the early nineties, when the COBE experiment (and then more recently WMAP [16]) measured for the first time fluctuations in the temperature CMBR at large angular separations, and found results consistent with the predictions of models with a HZ spectrum at such scales, the HZ type spectra have become a central pillar of standard models of structure formation in the universe.

Because of the highly irregular nature of structure at small scales, standard models with super-homogeneous features cannot be used even at zeroth order to describe these observed structures. This does not mean, however, that these models cannot describe successfully galaxy structures: but to establish whether they can, it must first be shown from observations that there is a clear crossover toward homogeneity i.e. a scale beyond which the average density becomes a well-defined (i.e. sample-independent) positive quantity. These models then predict that, on much larger scales (e.g.  $> 100$  Mpc), galaxy structures should present the super-homogeneous character of the HZ type PS. Indeed this should in principle be a critical test of the paradigm linking the measurements of CMBR on large scales to the distribution of matter. Observationally a crucial question is the feasibility of measuring the transition between these regimes directly in galaxy distributions. In this context one has to consider an important element: the galaxy distribution is a discrete set of objects whose properties are related in a non-trivial way to the ones of the underlying continuous field. To understand the relation between the two, one has to consider the additional effects related to sampling the continuous field. This is intimately related to the problem of “biasing” between the distribution of visible and dark matter which, as above mentioned, is usually invoked to explain a posteriori the observational features of  $\xi(r)$ .

Sampling a super-homogeneous fluctuation field may change the nature of correlations [17]. The reason can be found in the property of super-homogeneity of such distribution: the sampling, as for instance in the so-called “bias model” (selection of highest peaks of the fluctuations field) necessarily destroys the surface nature of the fluctuations, as it introduces a volume (Poisson-like) term in the variance. The “primordial” form of the power spectrum is thus not apparent in that which one would expect to measure from

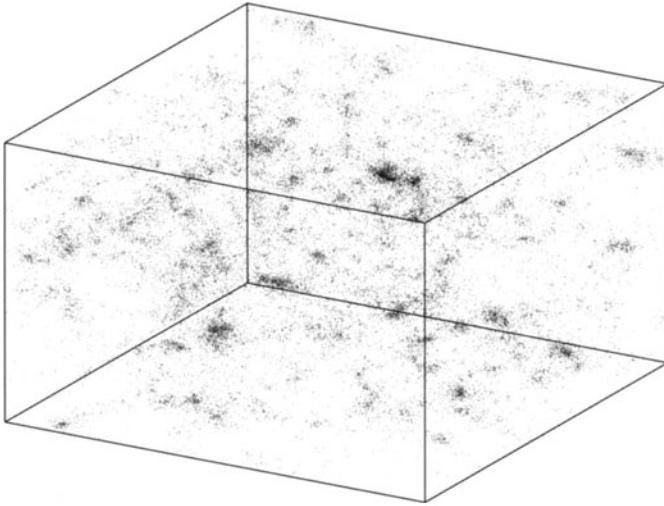


**Fig. 3.** In this figure we show a typical CDM type power spectrum (solid line) and the range associated to the primary observational constraints. The left hand side, the Harrison Zeldovich part of the PS ( $P(k) \sim k$ ), is constrained by observations of the anisotropies of the CMBR (dash-dotted box). Current galaxy and galaxy cluster surveys gives constraints at smaller scales (dashed box). The normalization of the amplitude of the galaxy or cluster PS to the one observed in the CMBR is fundamentally important. It is usually determined by a *linear* re-scaling on the y-axis, ascribed to the effect of bias. This simple assumption is not consistent with the canonical model for the biasing of a Gaussian field, which introduces a non-linear distortion both at small and large wave-number. This is illustrated by the dashed line, which shows what is actually obtained for the PS of the biased field. On small scales (large  $k$ ) there is a non-linear distortion and at large scales (small  $k$ ) the behavior is typical of a substantially Poisson system with  $P(k) \sim \text{const}$  (from [4])

objects selected in this way. This conclusion should hold for any generic model of bias [17]. If a linear amplification is obtained in some regime of scales (as it can be in certain phenomenological models of bias) it is necessarily a result of a fine-tuning of the model parameters. The study of different samplings and of the correlation features invariant under sampling represents an important issue in relation to the comparison of observations of galaxy structures, or distributions given by N-body simulations with primordial fluctuations (CMBR anisotropies) and theoretical models (see Fig. 3).

## 4 The Problem of Gravitational Structure Formation

In cosmology the main instrument for treating the *theoretical* problem of gravitational clustering is numerical (in the form of N-body simulations —NBS—), and the analytic understanding of this crucial problem is very limited. Other than in the regime of very small fluctuations where a linear analysis can



**Fig. 4.** This is a distribution obtained by starting from a Poisson particle configuration with zero initial velocity dispersion and by evolving it with gravity and by considering periodic boundary conditions. One may see that non-linear strong clustering is formed and its explanation is the central task of theoretical modeling. (Courtesy of T. Baertschiger)

be performed, the available models of clustering are essentially phenomenological models with numerous parameters which are fixed by numerical simulation. An NBS has these main features: (i) given  $N$  particles in a volume  $V$ , each particle moves according to the force given by the sum of the other  $N - 1$  particles and of all their replicas (due to the use of periodic boundary conditions). (ii) The equation of motions are integrated by using a leap-frog algorithm. (iii) Suitable techniques allow an appropriate summations (which makes the code faster) of far away particles contributions to the gravitational force. (iv) One may, or may not, consider the presence of space expansion. A typical result of the non-linear evolution is shown in Fig. 4.

A central issue in the context of cosmological NBS is to relate the formation of non linear structures to the specific choice of initial conditions used: this is done in order to constrain models with the observations of CMBR anisotropies, which are related to the initial conditions, and of galaxy structures, which give instead the final configuration of clustered matter. Standard primordial cosmological theoretical density fields, like the CDM case, are Gaussian and made of a huge number of very small mass particles, which are usually treated theoretically as a self-gravitating collision-less fluid: this means that the fluid must be dissipation-less and that two-body scattering should be negligible. The problem then being in which limit NBS, based on particle dynamics, are able to reproduce the two above conditions [7,18]. In this context one has to consider the issue of the physical role of particle fluc-

tuations in the dynamics of NBS. In fact, in the discretization of a continuous density field one faces two important limitations corresponding to the new length scales which are introduced. One the one hand a relatively small number of particles are used. This introduces a mass scale which is the mass of these particles. In typical cosmological NBS, this mass is of the order of a galaxy and hence many orders of magnitude larger than the microscopic mass of a CDM particle. Furthermore, it introduces a new length scale given by the average distance between nearest neighbor particles. On the other hand one must regularize the gravitational force at small scales in order to avoid numerical problems and typical small scale effects due to the discrete nature of the particles: given that the smoothening length is typically smaller than the initial inter-particle separation, it is not evident that one is effectively reproducing a dynamics where particles play the role of collision-less fluid elements. It is in this sense that one talks about the role of discreteness in NBS: that strong scattering between nearby particles are produced by the discretization and they should be considered artificial and spurious with respect to the dynamical evolution of a self-gravitating fluid. This point has been considered by many authors and they all show that discreteness has some influence on the formation of the structures [19]. Indeed, discreteness may play an important role in the early times formation of non-linear structures. How discrete effects are then “exported” toward large scales, if they are at all, is then an important but difficult problem to be understood. In other words the problem is that of understanding whether large non-linear structures, which at late times contain many particles, are produced solely by collision-less fluid dynamics, or whether the particle collisional processes are important also in the long-term, or whether they are made by a mix of these two effects.

For example in [8] we have presented an analysis of different sets of gravitational NBS all describing the dynamics of discrete particles with a small initial velocity dispersion. They encompass very different initial particle configurations, different numerical algorithms for the computation of the force, with or without the space expansion of cosmological models. Despite these differences we find in all cases that the non-linear clustering which results is essentially the same, with a well-defined simple power-law behavior in the conditional density in the range from a few times the lower cut-off in the gravitational force to the scale at which fluctuations are of order one. We have argued, presenting quantitative evidence, that this apparently universal behavior can be understood by the domination of the small scale contribution to the gravitational force, coming initially from nearest neighbor particles. A more quantitative description of this dynamics is evidently needed, with the principal goal of understanding the specific value observed of the exponent. In the cosmological literature (see e.g. [10]) the idea is widely dispersed that the exponents in non-linear clustering are related to that of the initial PS of the small fluctuations in the CDM fluid, and even that the non-linear

two-point correlation can be considered an analytic function of the initial two-point correlations. The models used to explain the behavior in the non-linear regime usually involve both the expansion of the Universe, and a description of the clustering in terms of the evolution of a continuous fluid. We have argued that the exponent is universal in a very wide sense, being common to the non-linear clustering observed in the non-expanding case. It would appear that the framework for understanding the non-linear clustering must be one in which discreteness (and hence intrinsically non-analytical behavior of the density field) is central, and that the simple context of non-expanding models should be sufficient to elucidate the essential physics.

## 5 Conclusion

There are then some basic questions which remain unanswered: does gravitational dynamics give rise to fractal clustering ? And, in case, from what initial conditions ? We believe there is much unexplored space for a statistical physics approach to the problem of gravitational clustering, which should be able to shed light on the more general characteristics of gravitational clustering.

## Acknowledgements

We warmly thanks our collaborators and particularly T. Baertschiger, A. Gabrielli and M. Joyce. FSL acknowledges the support of a Marie Curie Fellowship HPMF-CT-2001-01443.

## References

1. P.H. Coleman, L. Pietronero, Phys. Rep. **231**, 311, (1992).
2. F. Sylos Labini, M. Montuori, L. Pietronero, Phys. Rep. **293**, 66 (1998).
3. M. Joyce, M. Montuori, F. Sylos Labini, Astrophys. J. **514**, L5 (1999).
4. A. Gabrielli, F. Sylos Labini, M. Joyce, L. Pietronero, *Statistical Physics for Cosmic Structures* (Springer Verlag, Berlin, 2004).
5. A. Gabrielli, B. Jancovici, M. Joyce, J.L. Lebowitz, L. Pietronero, F. Sylos Labini, Phys. Rev. D **67**, 043406 (2003).
6. A. Gabrielli, M. Joyce, F. Sylos Labini, Phys. Rev. D **65**, 083523 (2002).
7. T. Baertschiger, M. Joyce, F. Sylos Labini, Astrophys. J. Lett. **581**, L63 (2002).
8. F. Sylos Labini, T. Baertschiger, M. Joyce: Europhys.Lett., **66**, 171 (2004).
9. K.K. Wu, O. Lahav, M. Rees, Nature **225**, 230, (1999).
10. P.J.E. Peebles, *Principles Of Physical Cosmology* (Princeton University Press, Princeton, New Jersey, 1993).
11. D. York et al., Astron. J. **120**, 1579 (2000).
12. B.B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1983).
13. I. Zehavi, et al., Astrophys. J. **571**, 172 (2002).

14. M. Joyce, P.W. Anderson, M. Montuori, L. Pietronero, F. Sylos Labini, Europhys. Lett. **50**, 416, (2000).
15. Yu.V. Baryshev, P. Teerikorpi , *The Discovery of Cosmic Fractals*, (World Scientific, Singapore, 2002).
16. C.L. Bennett, et al., Astrophys. J. Suppl. **148**, 1, (2003).
17. R. Durrer, A. Gabrielli, M. Joyce, F. Sylos Labini, Astrophys. J. Lett **585**, L1 (2003).
18. T. Baertschiger, F. Sylos Labini, Phys.Rev. D, **69**, 123001 (2004) [astro-ph/0401238](#).
19. M. Bottaccio, R. Capuzzo-Dolcetta, P. Miocchi, M. Montuori, L. Pietronero, Europhys. Lett. **7**, 315 (2002).

# Local Scale Invariance in Ageing Phenomena

Malte Henkel

Laboratoire de Physique des Matériaux (CNRS UMR 7556)  
Université Henri Poincaré Nancy I  
B.P. 239, 54506 Vandœuvre lès Nancy Cedex, France

**Abstract.** Many materials quenched into their ordered phase undergo ageing and they show dynamical scaling. For any given dynamical exponent  $z$ , this can be extended to a new form of local scale invariance which acts as a dynamical symmetry. The scaling functions of the two-time correlation and response functions of ferromagnets with a non-conserved order parameter are determined. These results are in agreement with analytical and numerical studies of various models, especially the kinetic Glauber-Ising model in 2 and 3 dimensions.

## 1 Introduction

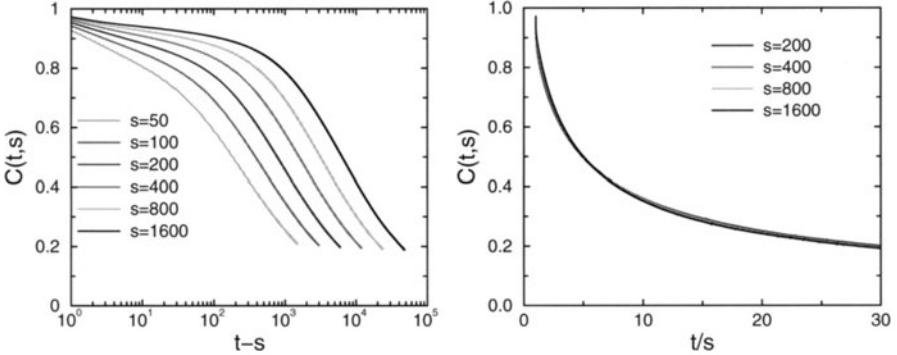
Ageing in its most general sense refers to the change of material properties as a function of time. In particular, *physical ageing* occurs when the underlying microscopic processes are reversible while on the other hand, biological systems age because of irreversible chemical reactions going on within them. Historically, ageing phenomena were first observed in glassy systems, see [1], but it is of interest to study them in systems without disorder. These should be conceptually simpler and therefore allow for a better understanding. Insights gained this way may become useful for a later study of glassy systems.

## 2 Phenomenology of Ageing

In describing the phenomenology of ageing system, we shall refer throughout to simple ferromagnets, see [2,3,4,5] for reviews. We consider systems which undergo a second-order equilibrium phase transition at a critical temperature  $T_c > 0$  and we shall assume throughout that the dynamics admits no macroscopic conservation law. Initially, the system is prepared in some initial state (typically one considers an initial temperature  $T_{\text{ini}} = \infty$ ). The system is brought out of equilibrium by quenching it to a final temperature  $T \leq T_c$ . Then  $T$  is fixed and the system's temporal evolution is studied. It turns out that the relaxation back to global equilibrium is very slow (e.g. algebraic in time) with a formally infinite relaxation time for *all*  $T \leq T_c$ .

Let  $\phi(t, \mathbf{r})$  denote the time- and space-dependent order parameter and consider the two-time correlation and (linear) response functions

$$C(t, s; \mathbf{r}) = \langle \phi(t, \mathbf{r}) \phi(s, \mathbf{0}) \rangle , \quad R(t, s; \mathbf{r}) = \left. \frac{\delta \langle \phi(t, \mathbf{r}) \rangle}{\delta h(s, \mathbf{0})} \right|_{h=0} \quad (1)$$



**Fig. 1.** Two-time autocorrelator in the 2D Glauber-Ising model, with temperature  $T = 1.5 = 0.66T_c$  on a  $600^2$  lattice and a disordered initial state

where  $h$  is the magnetic field conjugate to  $\phi$  and space-translation invariance was already implicitly assumed. The autocorrelation and autoreponse functions are given by  $C(t, s) = C(t, s; \mathbf{0})$  and  $R(t, s) = R(t, s; \mathbf{0})$  where  $t$  is referred to as *observation time* and  $s$  is called the *waiting time*.

In Fig. 1 we show the autocorrelator  $C(t, s)$  of the 2D kinetic Ising model with Glauber dynamics after a quench to the final temperature  $T = 1.5$ . In the left panel, the dependence of  $C(t, s)$  on the time difference  $\tau = t - s$  is shown. Clearly, the autocorrelator depends on both  $t$  and  $s$ . For large values of  $s$  and  $\tau \lesssim s$ , the values of  $C(t, s)$  reach a quasistationary value  $C_{\text{qs}}(\tau) \simeq M_{\text{eq}}^2$ , where  $M_{\text{eq}}$  is the equilibrium magnetization. In the regime  $\tau \gtrsim s$  one observes an algebraic decay of  $C(t, s)$ . Qualitatively similar behavior is known from glassy systems and the simultaneous dependence of  $C(t, s)$  and/or  $R(t, s)$  on *both*  $t$  and  $s$  is the formal definition of *ageing behaviour*. The strong dependence of  $C(t, s)$  on the waiting time (which expresses the sensibility of the system's properties on its entire history) seems at first sight to lead to irreproducible data and hence to prevent a theoretical understanding of the ageing phenomenon. Remarkably, Struik [1] observed in polymeric glasses submitted to mechanical stress that the linear responses of quite distinct materials could be mapped onto a single and *universal* master curve. We illustrate this here in the ferromagnetic Glauber-Ising model through the data collapse in the right panel of Fig. 1. Remarkably, a dynamical scaling holds although the equilibrium state need not be scale-invariant.

On a more microscopic level, correlated domains of a linear size  $L(t)$  form. These are ordered if  $T < T_c$  but do contain internal long-range fluctuations at criticality. In the first case, the system undergoes *phase-ordering kinetics* and in the second *non-equilibrium critical dynamics*. For sufficiently large times, the domain size scales with time as  $L(t) \sim t^{1/z}$  where  $z$  is the dynamical exponent. The slow relaxation to global equilibrium (although local equilibrium is rapidly achieved) comes about since for  $T < T_c$  there are at least two distinct and competing equilibrium states. These states merge at

**Table 1.** Values of the non-equilibrium exponents  $a$ ,  $b$  and  $z$  for non-conserved ferromagnets with  $T_c > 0$ . The non-trivial critical-point value  $z_c$  is model-dependent

	$a$	$b$	$z$	Class
$T = T_c$	$(d - 2 + \eta)/z$	$(d - 2 + \eta)/z$	$z_c$	L
$T < T_c$	$(d - 2 + \eta)/z$	0	2	L
	$1/z$	0	2	S

$T = T_c$ . On each site  $\mathbf{r}$  the local environment selects the local equilibrium state.

As suggested from Fig. 1, one expects a scaling regime to occur when

$$t \gg \tau_{\text{micro}} , s \gg \tau_{\text{micro}} , t - s \gg \tau_{\text{micro}} \quad (2)$$

where  $\tau_{\text{micro}}$  is some ‘microscopic’ time scale. We shall see later how important the third condition in (2) is. If the conditions (2) hold, one expects [2,4]

$$C(t, s) = s^{-b} f_C(t/s) , f_C(y) \sim y^{-\lambda_C/z} ; y \rightarrow \infty \quad (3a)$$

$$R(t, s) = s^{-1-a} f_R(t/s) , f_R(y) \sim y^{-\lambda_R/z} ; y \rightarrow \infty \quad (3b)$$

These scaling forms should hold for both  $T < T_c$  and  $T = T_c$  although the values of the exponents will in general be different in these two cases. Here  $\lambda_C$  and  $\lambda_R$  are the autocorrelation [6] and autoresponse [7] exponents, respectively. They are independent of the equilibrium exponents and of  $z$  [8]. It was taken for granted since a long time that  $\lambda_C = \lambda_R$  but examples to the contrary have recently been found for spatial long-range correlations in the initial data [7] and in the random-phase sine-Gordon model [9]. If  $T_{\text{ini}} = \infty$ , the inequality  $\lambda_C = \lambda_R \geq d/2$  holds [10].

The values of the non-equilibrium exponents  $a$  and  $b$  apparently depend on properties of the equilibrium system as follows [11] and are listed in Tab. 1, together with those of  $z$ . We restrict to non-conserved ferromagnetic systems with  $T_c > 0$ . If the equilibrium order parameter correlator  $C_{\text{eq}}(\mathbf{r}) \sim \exp(-|\mathbf{r}|/\xi)$  with a finite  $\xi$ , the system is said to be of *class S* and if  $C_{\text{eq}}(\mathbf{r}) \sim |\mathbf{r}|^{-(d-2+\eta)}$ , it is said to be of *class L*. At criticality, a system is always in class L, but if  $T < T_c$ , systems such as the Glauber-Ising model are in class S, whereas the kinetic spherical model is in class L. For class S, the value of  $a$  comes from the well-accepted idea [12,3] that the time-dependence of macroscopic averages comes from the motion of the domain walls. For class L, it follows from a hyperscaling argument [13].

Having fixed the values of the critical exponents, we can state our main question: what can be said on the form of the universal scaling functions  $f_C(y)$ ,  $f_R(y)$  in a general, model-independent way ?

### 3 Local Scale Invariance

Our starting point is the rich evidence, accumulated through many decades and reviewed in [2], in favour of dynamical scale invariance in ageing phenomena. The order parameter field  $\phi = \phi(t, \mathbf{r})$  scales

$$\phi(t, \mathbf{r}) = b^{-x_\phi} \phi(b^{-z} t, b^{-1} \mathbf{r}) \quad (4)$$

where  $b$  is a constant rescaling factor. We now ask whether eq. (4) can be sensibly generalized to general space-time dependent rescaling  $b = b(t, \mathbf{r})$  [14,15,16]. This ansatz can be motivated as follows.

**Example 1.** Consider an equilibrium critical point in  $(1+1)$ -dimensional space-time. Then  $z = 1$  and let  $w = t + ir$  and  $\bar{w} = t - ir$ . Any angle-preserving space-time transformation is conformal and is given by the analytic transformations  $w \mapsto f(w), \bar{w} \mapsto \bar{f}(\bar{w})$ . A well-known result from field theory states [17] that for short-ranged interactions, there is a Ward identity such that invariance under space- and time-translations, rotations and dilatations implies conformal invariance. Furthermore, basic quantities as the order parameter are *primary* under the conformal group and transform as  $\phi(w, \bar{w}) \mapsto (f'(w)\bar{f}'(\bar{w}))^{x_\phi/2} \phi(f(w)\bar{f}(\bar{w}))$  [18]. Hence  $n$ -point correlation functions and the values of the exponents  $x_\phi$  can be found exactly from conformal symmetry, see e.g. [17,19] for introductions. Here we merely need the projective conformal transformations  $f(w) = (\alpha w + \beta)/(\gamma w + \delta)$  with  $\alpha\delta - \beta\gamma = 1$ . Fields which transform covariantly under those are called *quasiprimary* [18]. The associated infinitesimal transformations are  $\ell_n = -w^{n+1}\partial_w$ ,  $n \in \{\pm 1, 0\}$ , and satisfy the Lie algebra  $[\ell_n, \ell_{n'}] = (n - n')\ell_{n+n'}$ .

**Example 2.** Let  $z = 2$  and consider  $d$  space dimensions. The *Schrödinger group*  $Sch(d)$  is defined by [20]

$$t \mapsto \frac{\alpha t + \beta}{\gamma t + \delta}, \quad \mathbf{r} \mapsto \frac{\mathcal{R}\mathbf{r} + \mathbf{v}t + \mathbf{a}}{\gamma t + \delta}; \quad \alpha\delta - \beta\gamma = 1 \quad (5)$$

where  $\mathcal{R} \in SO(d)$ ,  $\mathbf{a}, \mathbf{v} \in \mathbb{R}^d$  and  $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ . It is well-known that  $Sch(d)$  is the maximal kinematic group of the free Schrödinger equation  $\mathcal{S}\psi = 0$  with  $\mathcal{S} = 2mi\partial_t - \partial_{\mathbf{r}}^2$  [20] (that is, it maps any solution of  $\mathcal{S}\psi = 0$  to another solution). There are many Schrödinger-invariant systems, e.g. non-relativistic free fields [21] or the Euler equations of fluid dynamics [22]. As in the conformal case, for local theories there is a Ward identity such that [23]

$$\left. \begin{array}{l} \text{space translation invariance} \\ \text{scale invariance with } z = 2 \\ \text{Galilei invariance} \end{array} \right\} \implies \text{Schrödinger invariance} \quad (6)$$

We point out that Galilei invariance has to be required and for applications to ageing we note that time-translation invariance is not needed. Indeed, a non-trivial Galilei invariance is only possible for a complex wave function  $\psi$ .

In applications to ageing, we shall identify below the ‘complex conjugate’ of the order parameter  $\phi^* = \tilde{\phi}$  with the response field of non-equilibrium field theory [23]. We denote the Lie algebra of  $Sch(d)$  by  $\mathfrak{sch}_d$ . Specifically,  $\mathfrak{sch}_1 = \{X_{\pm 1,0}, Y_{\pm 1/2}, M_0\}$  with the non-vanishing commutation relations

$$[X_n, X_{n'}] = (n - n')X_{n+n'}, [X_n, Y_m] = \left(\frac{n}{2} - m\right)Y_{n+m}, [Y_{1/2}, Y_{-1/2}] = M_0 \quad (7)$$

where  $n, n' \in \{\pm 1, 0\}$  and  $m \in \{\pm 1/2\}$ .

**Example 3.** For a dynamical exponent  $z \neq 2$ , we construct infinitesimal generators of local scale transformations from the following requirements [16] (for simplicity, set  $d = 1$ ): (a) Transformations in time are  $t \mapsto (at + \beta)/(\gamma t + \delta)$  with  $\alpha\delta - \beta\gamma = 1$ . (b) The generator for time-translations is  $X_{-1} = -\partial_t$  and for dilatations  $X_0 = -t\partial_t - z^{-1}r\partial_r - x/z$ , where  $x$  is the scaling dimension of the fields  $\phi, \tilde{\phi}$  on which the generators act. (c) Space-translation invariance is required, with generator  $-\partial_r$ . Starting from these conditions, we can show by explicit construction that there exist generators  $X_n$ ,  $n \in \{\pm 1, 0\}$ , and  $Y_m$ ,  $m = -1/z, 1 - 1/z, \dots$  such that

$$[X_n, X_{n'}] = (n - n')X_{n+n'}, [X_n, Y_m] = \left(\frac{n}{z} - m\right)Y_{n+m} \quad (8)$$

For generic values of  $z$ , it is sufficient to specify the ‘special’ generator [16]

$$X_1 = -t^2\partial_t - Ntr\partial_r - Nxt - \tilde{\alpha}r^2\partial_t^{N-1} - \tilde{\beta}r^2\partial_r^{2(N-1)/N} - \tilde{\gamma}\partial_r^{2(N-1)/N}r^2 \quad (9)$$

from which all other generators can be recovered and where we wrote  $z = 2/N$  and  $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}$  are free constants (further non-generic solutions exist for  $N = 1$  and  $N = 2$ ). For  $z = 2$  we recover the Schrödinger Lie algebra  $\mathfrak{sch}_1$ . Now, the condition  $[X_1, Y_{N/2}] = 0$  is only satisfied if either (I)  $\tilde{\beta} = \tilde{\gamma} = 0$  which we call *type I* or else (II)  $\tilde{\alpha} = 0$  which we call *type II* [16].

**Definition:** If a system is invariant under the generators of either type I or type II it is said to be *locally scale-invariant* of type I or type II, respectively.

Local scale invariance of type I can be used to describe strongly anisotropic equilibrium critical points. The application to Lifshitz points in 3D magnets with competing interactions is discussed in [24,16]. The generators of type II are suitable for applications to ageing phenomena and will be studied here. First, we note that the generators  $X_n, Y_m$  form a kinematic symmetry of the linear differential equation  $\mathcal{S}\psi = 0$  where  $\mathcal{S} = -z^2(\tilde{\beta} + \tilde{\gamma})\partial_t + \partial_r^z$  [16]. Recently, systems of non-linear equations invariant under these generators with  $\tilde{\alpha} = \tilde{\beta} = \tilde{\gamma} = 0$  but extended to an infinite-dimensional symmetry  $t \mapsto f(t)$  have been found [25]. Second, we consider the consequences for the scaling form of the response function  $R(t, s; \mathbf{r})$ . To do this, we recall that in the context of Martin-Siggia-Rose theory (see [26]) a response function  $R(t, s) = \langle \phi(t)\tilde{\phi}(s) \rangle$  may be viewed as a correlator. If

both  $\phi$  and  $\tilde{\phi}$  transform as quasiprimaries, the hypothesis of covariance of the autoresponse function leads to the two conditions  $X_0 R = X_1 R = 0$ . Of course, ageing systems cannot be invariant under time-translations. From the explicit form of the generators given above these equations are easily solved and the result can be compared with the expected asymptotic behavior (3). This leads to the general result [27,16]

$$R(t, s) = r_0 \Theta(t - s) \left( \frac{t}{s} \right)^{1+a-\lambda_R/z} (t - s)^{-1-a} \quad (10)$$

where  $r_0$  is a normalization constant and the causality condition  $t > s$  is explicitly included. Furthermore, the space-time response is given by  $R(t, s; \mathbf{r}) = R(t, s) \Phi(r(t - s)^{-1/z})$  where  $\Phi(u)$  solves the equation [16]

$$\left[ \partial_u + z \left( \tilde{\beta} + \tilde{\gamma} \right) u \partial_u^{2-z} + 2z(2-z)\tilde{\gamma} \partial_u^{1-z} \right] \Phi(u) = 0 \quad (11)$$

In the special case  $z = 2$ , this reduces to [14]

$$R(t, s; \mathbf{r}) = R(t, s) \exp \left( -\frac{\mathcal{M}}{2} \frac{\mathbf{r}^2}{t - s} \right) \quad (12)$$

where  $\mathcal{M} = \tilde{\beta} + \tilde{\gamma}$  is constant.

We point out that the derivation of the space-time response needs the assumption of Galilei invariance (suitably generalized if  $z \neq 2$ ). In turn, the confirmation of the form (12) is a given system undergoing ageing provides evidence in favour of Galilei invariance in that system. We shall next describe tests of (10,12) in the Glauber-Ising model in  $d \geq 2$  dimensions before we return to a fuller discussion of the physical origins of local scale invariance.

## 4 Numerical Test in the Glauber-Ising Model

We wish to test the predictions (10,12) of local scale invariance in the kinetic Glauber-Ising model, defined by the Hamiltonian  $\mathcal{H} = -\sum_{(\mathbf{i}, \mathbf{j})} \sigma_{\mathbf{i}} \sigma_{\mathbf{j}}$  where  $\sigma_{\mathbf{i}} = \pm 1$ . Based on a master equation, we use the heat-bath stochastic rule

$$\sigma_{\mathbf{i}}(t+1) = \pm 1 \text{ with probability } \frac{1}{2} [1 \pm \tanh(h_{\mathbf{i}}(t)/T)] \quad (13)$$

with the local field  $h_{\mathbf{i}}(t) = \sum_{\mathbf{n}(\mathbf{i})} \sigma_{\mathbf{n}}(t)$  and  $\mathbf{n}(\mathbf{i})$  runs over the nearest neighbors of the site  $\mathbf{i}$ .

The response function is too noisy to be measured directly, therefore following [28] one may add a quenched spatially random magnetic field  $\pm h_{(0)}$  between the times  $t_1$  and  $t_2$  and measure the integrated response  $M(t, t_1, t_2) := h_{(0)} \int_{t_1}^{t_2} du R(t, u)$ . Two schemes are widely used, namely the ‘zero-field-cooling’ (ZFC) scheme, where  $t_1 = s$  and  $t_2 = t$  and the ‘thermoremanent’ (TRM) scheme, where  $t_1 = 0$  and  $t_2 = s$ . However, in both schemes it is not possible

**Table 2.** Values of the autoresponse exponent  $\lambda_R$  and of the parameters  $r_0$ ,  $r_1$  and  $\mathcal{M}$  in the Glauber-Ising model for an infinite-temperature initial state

$d$	$T$	$\lambda_R$	$r_0$	$r_1$	$\mathcal{M}$
2	1.5	1.26	$1.76 \pm 0.03$	$-1.84 \pm 0.03$	$4.08 \pm 0.04$
3	3	1.60	$0.10 \pm 0.01$	$0.20 \pm 0.01$	$4.22 \pm 0.05$

to naïvely use the scaling form (3) and integrate in order to obtain  $M$ . This comes about since in both cases some of the conditions (2) for the validity of this scaling form are violated. Taking this fact into account leads to the following results [11,13]: (a) the thermoremanent magnetization

$$\rho(t, s) := \int_0^s du R(t, u) = r_0 s^{-a} f_M(t/s) + r_1 s^{-\lambda_R/z} g_M(t/s) \quad (14)$$

$$f_M(y) = y^{-\lambda_R/z} {}_2F_1 \left( 1 + a, \frac{\lambda_R}{z} - a; \frac{\lambda_R}{z} - a + 1; \frac{1}{y} \right), \quad g_M(y) \simeq y^{-\lambda_R/z}$$

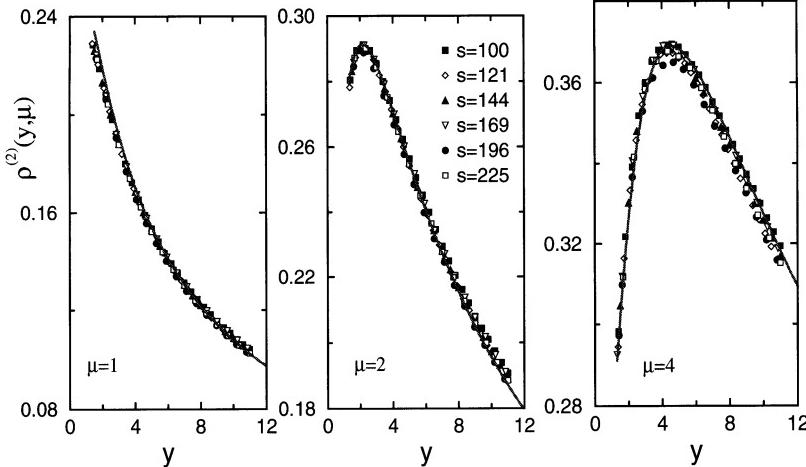
where  $r_{0,1}$  are normalization constants. The first term is as expected from naïve scaling. In practice,  $a$  and  $\lambda_R/z$  are often quite close and the size of the correction term may well be notable for  $T < T_c$  (at  $T = T_c$ ,  $a$  and  $\lambda_R/z$  are usually quite distinct); (b) the zero-field-cooled susceptibility

$$\chi(t, s) := \int_s^t du R(t, u) = \chi_0 + s^{-A} g(t/s) + O(s^{-a}) \quad (15)$$

with a constant  $\chi_0$  and some scaling function  $g$ . For systems of class S, we have  $A = a - \kappa$ , where  $\kappa$  measures the width  $w(t) \sim t^\kappa$  of the domain walls [13]. In the Glauber-Ising model, one has  $\kappa = 1/4$  in 2D and  $w(t) \sim \sqrt{\ln t}$  in 3D [29], while  $\kappa = 0$  for  $d > 3$ . Consequently, the term of order  $s^{-a}$  coming from naïve scaling is not even the dominant one in the long-time limit  $s \rightarrow \infty$  and a simple phenomenological analysis of data of  $\chi(t, s)$  is likely to produce misleading results. For systems of class L,  $A = 0$ .

Indeed, based on high-quality numerical MC data for  $\chi(t, s)$  in the 2D Glauber-Ising model and performing a straightforward scaling analysis according to  $\chi(t, s) \sim s^{-a}$  but *without* taking the third condition (2) for the validity of scaling into account, it had been claimed that  $a = 1/4$  in that model [30]. However, that analysis is based on the identification  $A = ? a$  which cannot be maintained. Rather, for the 2D Glauber-Ising model, one has  $a = 1/2$  and  $\kappa = 1/4$ , reproducing  $A = 1/4$  in agreement with the MC data.

After these preparations, we can now present numerical Monte Carlo (MC) data and compare them with the predictions (10,12). We consider the thermoremanent magnetization  $\rho(t, s)$  and subtract off the leading finite-time



**Fig. 2.** Integrated space-time response of the 2D Glauber-Ising model. After [31]

correction according to (14). For  $T < T_c$  this leads to the parameter values collected in Tab. 2, see [31] for details. Then the MC data both at  $T = T_c$  and for  $T < T_c$  for  $\rho(t, s)$  are in full agreement with (14) [27,31], in both 2D and 3D. Here we present a direct test of Galilei invariance by considering the space-time integrated response

$$\frac{d\rho(t, s; \mu)}{d\Omega} = T \int_0^s du \int_0^{\sqrt{\mu s}} dr r^{d-1} R(t, u; \mathbf{r}) = r_0 s^{d/2-a} \rho^{(2)}(t/s, \mu) \quad (16)$$

with an explicitly known expression for  $\rho^{(2)}$  following from (12) and the leading finite-time correction is already subtracted off [31]. Since all non-universal parameters were determined before and are listed in Tab. 2, this comparison between simulation and local scale invariance is parameter-free. The result is shown in Fig. 2 in 2D and we find a perfect agreement. A similar results holds in 3D [31].

This direct evidence in favour of Galilei invariance in the phase-ordering kinetics of the Glauber-Ising model is all the more remarkable since the zero-temperature time-dependent Ginzburg-Landau equation (TDGL), which is usually thought to describe the same system (e.g. [2]), does not have this symmetry. Indeed a recent second-order result for  $R(t, s)$  does not agree with (10) [32] (similar corrections also arise at  $T = T_c$  [33]). On the other hand,  $\lambda_C = 1$  from the exact solution of the 1D Glauber-Ising model at  $T = 0$  [4], while  $\lambda_C \simeq 0.6006\dots$  in the 1D TDGL [34], implying that these two models belong to distinct universality classes.

Confirmations of (10,12) in exactly solvable models are reviewed in [16].

## 5 Influence of Noise

We now wish to review the present state of theoretical arguments [23,35] in order to understand from where the recent numerical evidence in favour of a larger dynamical symmetry than mere scale invariance in ageing phenomena might come from. We shall do this here for phase-ordering kinetics. Then  $z = 2$  and we have to consider the Schrödinger group and Schrödinger-invariant systems. For simplicity, we often set  $d = 1$ . From the following discussion, the importance of Galilei invariance will become clear, see also (6).

**A)** Consider the free Schrödinger equation  $(2\mathcal{M}\partial_t - \partial_r^2)\phi = 0$  where  $\mathcal{M} = im$  is fixed. While an element of the Schrödinger group acts projectively (i.e. up to a known companion function [20]) on the wave function  $\phi$ , we can go over to a true representation by treating  $\mathcal{M}$  as an additional variable. Following [36], we define a new coordinate  $\zeta$  and a new wave function  $\psi$  by

$$\phi(t, r) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} d\zeta e^{-i\mathcal{M}\zeta} \psi(\zeta, t, r) \quad (17)$$

We denote time  $t$  as the zeroth coordinate and  $\zeta$  as coordinate number  $-1$ .

We inquire about the maximal kinematic group in this case [23]. Now, the projective phase factors can be absorbed into certain translations of the variable  $\zeta$  [23]. Furthermore, the free Schrödinger equation becomes

$$(2i\partial_\zeta\partial_t + \partial_r^2) \psi(\zeta, t, r) = 0 \quad (18)$$

In order to find the maximal kinematic symmetry of this equation, we recall that the three-dimensional Klein-Gordon equation  $\sum_{\mu=-1}^1 \partial_\mu \partial^\mu \Psi(\xi) = 0$  has the 3D conformal algebra  $\mathbf{conf}_3 \cong so(4, 1)$  as maximal kinematic symmetry. By making the following change of variables

$$\zeta = (\xi_0 + i\xi_{-1})/2, \quad t = (-\xi_0 + i\xi_{-1})/2, \quad r = \xi_1 \sqrt{1/2} \quad (19)$$

and setting  $\psi(\zeta, t, r) = \Psi(\xi)$ , the 3D Klein-Gordon equation reduces to (18). Therefore, for variable masses  $\mathcal{M}$ , the maximal kinematic symmetry algebra of the free Schrödinger equation in  $d$  dimensions is isomorphic to the conformal algebra  $\mathbf{conf}_{d+2}$  and we have the inclusion of the complexified Lie algebras  $(\mathfrak{sch}_d)_C \subset (\mathbf{conf}_{d+2})_C$  [23,37].

**B)** The Galilei invariance of the free Schrödinger equation requires the existence of a formal ‘complex conjugate’  $\phi^*$  of the order parameter  $\phi$ . On the other hand, a common starting point in the description of ageing phenomena is a Langevin equation which may be turned into a field theory using the Martin-Siggia-Rose (MSR) formalism [26,17] and which involves besides  $\phi$  the response field  $\tilde{\phi}$ . If we identify  $\phi^* = \tilde{\phi}$  and use (17) together with the assumption that  $\psi$  is real to define the complex conjugate, then the causality condition that

$$R(t, s; \mathbf{r}) = \langle \phi(t, \mathbf{r}) \tilde{\phi}(s, \mathbf{0}) \rangle = \langle \phi(t, \mathbf{r}) \phi^*(s, \mathbf{0}) \rangle \quad (20)$$

vanishes for  $t < s$ , follows naturally (and similarly for three-point response functions) [23]. Therefore, the calculation of response and of correlation functions from a dynamical symmetry should be done in the same way.

**C)** So far, we have concentrated exclusively in applications of local scale invariance to finding the form of response functions while the determination of correlation functions was not yet addressed. We shall do so now and consider the Langevin equation (with  $D^{-1} = 2\mathcal{M}$ ) [35]

$$\partial_t \phi = -D \frac{\delta \mathcal{H}}{\delta \phi} - Dv(t)\phi + \eta \quad (21)$$

where  $\mathcal{H}$  is the usual Ginzburg-Landau functional,  $v(t)$  is a time-dependent Lagrange multiplier which will be chosen to produce the constraint  $C(t, t) = 1$  and  $\eta$  is an uncorrelated Gaussian noise describing the coupling to a heat bath such that  $\langle \eta \rangle = 0$  and  $\langle \eta(t, \mathbf{r})\eta(t', \mathbf{r}') \rangle = 2DT\delta(t-t')\delta(\mathbf{r}-\mathbf{r}')$ . Another source of noise comes from the initial conditions and we shall always use an uncorrelated initial state such that

$$C(0, 0; \mathbf{r}) = \langle \phi(0, \mathbf{r})\phi(0, \mathbf{0}) \rangle = a_0\delta(\mathbf{r}) \quad (22)$$

where  $a_0$  is a constant.

The MSR action of (21) reads  $S[\phi, \tilde{\phi}] = S_0[\phi, \tilde{\phi}] + S_b[\phi, \tilde{\phi}]$  where

$$S_0[\phi, \tilde{\phi}] = \int dt d\mathbf{r} \tilde{\phi} \left( \frac{\partial \phi}{\partial t} + D \frac{\delta \mathcal{H}}{\delta \phi} + Dv(t)\phi \right) \quad (23a)$$

$$S_b[\phi, \tilde{\phi}] = -DT \int dt d\mathbf{r} \tilde{\phi}(t, \mathbf{r})^2 - \frac{a_0}{2} \int d\mathbf{r} \tilde{\phi}(0, \mathbf{r})^2 \quad (23b)$$

and we used (22), see [32]. Here  $S_0$  describes the noiseless part of the action while the thermal and the initial noise are contained in  $S_b$ . Finally, the potential  $v(t)$  can be absorbed into a gauge transformation; for example if  $\phi^{(0)}$  is a solution of the free Schrödinger equation, then  $\phi = \phi^{(0)}k(t)$  solves the Schrödinger equation with the potential  $v(t)$  and where

$$k(t) := \exp \left( -D \int^t du v(u) \right) \quad (24)$$

The realization of the Schrödinger algebra with  $v(t) \neq 0$  is easily found [35].

We now assume in addition to dynamical scaling that  $\mathcal{H}$  is such that *at temperature  $T = 0$ , the theory is Galilei invariant* [35]. This looks physically reasonable and we now explore some consequences of this hypothesis. We denote by  $\langle \cdot \rangle_0$  an average carried out using only the noiseless part  $S_0$  of the action. The Bargman superselection rules state that  $\underbrace{\langle \phi \cdots \phi}_{n} \underbrace{\tilde{\phi} \cdots \tilde{\phi} \rangle}_m_0 = 0$  if

$n \neq m$ . First, the response function is

$$\begin{aligned} R(t, s; \mathbf{r}) &= \langle \phi(t, \mathbf{r})\tilde{\phi}(s, \mathbf{0}) \rangle = \langle \phi(t, \mathbf{r})\tilde{\phi}(s, \mathbf{0}) \exp(-S_b[\phi, \tilde{\phi}]) \rangle_0 \\ &= \langle \phi(t, \mathbf{r})\tilde{\phi}(s, \mathbf{0}) \rangle_0 =: R_0(t, s; \mathbf{r}) \end{aligned} \quad (25)$$

where in the last line the exponential was expanded and the Bargman superselection rule was used. Here  $R_0$  is the noiseless response and the form (10,12) of Schrödinger invariance is recovered if  $v(t) = (2\mathcal{M})(1 + a - \lambda_R/2)t^{-1}$  [35]. In other words, under the stated hypothesis, the response function is independent of the noises. This is certainly in agreement with the explicit model calculations reviewed in section 3.

Second, we now obtain the autocorrelation function. As before [35]

$$\begin{aligned} C(t, s) &= \left\langle \phi(t, \mathbf{r})\phi(s, \mathbf{r}) \exp\left(-S_b[\phi, \tilde{\phi}]\right) \right\rangle_0 \\ &= DT \int du d\mathbf{R} R_0^{(3)}(t, s, u; \mathbf{R}) + \frac{a_0}{2} \int d\mathbf{R} R_0^{(3)}(t, s, 0; \mathbf{R}) \end{aligned} \quad (26)$$

where  $R_0^{(3)}(t, s, u; \mathbf{R}) = \langle \phi(t, \mathbf{y})\phi(s, \mathbf{y})\tilde{\phi}(u, \mathbf{R} + \mathbf{y})^2 \rangle_0$ . In contrast with the response function, the autocorrelation function contains *only* noisy terms and in fact vanishes in the absence of noise. By hypothesis, Schrödinger invariance holds for the noiseless theory and the three-point function  $R_0^{(3)}$  is fixed up to a scaling function of a single variable [14,35].

Working out the asymptotic behavior of  $C(t, s)$  for  $y = t/s \rightarrow \infty$  according to (3) and comparing with the response function (25), we find that *for any coarsening system with a disordered initial state (22) and whose noiseless part is Schrödinger-invariant, the relation  $\lambda_C = \lambda_R$  holds true* [35]. For the first time a general sufficient criterion for this exponent relation is found.

The autocorrelator scaling function becomes for phase-ordering

$$f_C(y) = \frac{a_0}{2} y^{\lambda_C/2} (y - 1)^{-\lambda_C} \Phi\left(\frac{y+1}{y-1}\right) \quad (27)$$

but the scaling function  $\Phi(w)$  is left undetermined by Schrödinger invariance. If in addition we require that  $C(t, s)$  should be non-singular at  $t = s$ , the asymptotic behavior  $\Phi(w) \sim w^{-\lambda_C}$  for  $w \rightarrow \infty$  follows. Provided that form should hold true for all values of  $w$ , we would obtain approximately

$$f_C(y) \approx f_0 \left( \frac{(y+1)^2}{4y} \right)^{-\lambda_C/2} \quad (28)$$

Indeed, this is found to be satisfied for several ageing spin systems with an underlying free-field theory [35]. On the other hand, (28) does not hold true in the Glauber-Ising model. Work is presently being carried out in order to describe  $f_C(y)$  in this model and will be reported elsewhere [38].

## Acknowledgements

It is a pleasure to thank M. Pleimling, A. Picone and J. Unterberger for the fruitful collaborations which led to the results reviewed here. This work was supported by CINES Montpellier (projet pmn2095) and by the Bayerisch-Französisches Hochschulzentrum (BFHZ).

## References

1. L.C.E. Struik *Physical ageing in amorphous polymers and other materials* (Elsevier, Amsterdam 1978).
2. A.J. Bray Adv. Phys. **43**, 357 (1994).
3. J.P. Bouchaud in M.E. Cates, M.R. Evans (eds) *Soft and fragile matter* (IOP Press, Bristol 2000).
4. C. Godrèche, J.M. Luck, J. Phys. Cond. Matt. **14**, 1589 (2002).
5. L.F. Cugliandolo, in: *Slow Relaxation and non equilibrium dynamics in condensed matter*, Les Houches Session 77 July 2002, J-L Barrat, J Dalibard, J Kurchan, M V Feigel'man eds (Springer, Heidelberg 2003).
6. D.S. Fisher, D.A. Huse, Phys. Rev. B **38**, 373 (1988).
7. A. Picone, M. Henkel, J. Phys. A **35**, 5575 (2002).
8. H.K. Janssen, B. Schaub, B. Schmittmann, Z. Phys. B **73**, 539 (1989).
9. G. Schehr, P. Le Doussal, Phys. Rev. E **68**, 046101 (2003).
10. C. Yeung, M. Rao, R.C. Desai, Phys. Rev. E **53**, 3073 (1996).
11. M. Henkel, M. Paefßens, M. Pleimling, Europhys. Lett. **62**, 644 (2003).
12. L. Berthier, J.L. Barrat, and J. Kurchan, Eur. Phys. J. B **11**, 635 (1999).
13. M. Henkel, M. Paefßens, M. Pleimling, Phys. Rev. E **69**, 056109 (2004).
14. M. Henkel, J. Stat. Phys. **75**, 1023 (1994).
15. M. Henkel, Phys. Rev. Lett. **78**, 1940 (1997).
16. M. Henkel, Nucl. Phys. B **641**, 405 (2002).
17. J.L. Cardy, *Scaling and Renormalization in Statistical Mechanics* (Cambridge University Press, Cambridge 1996).
18. A.A. Belavin, A.M. Polyakov, A.B. Zamolodchikov, Nucl. Phys. B **241**, 333 (1984).
19. M. Henkel, *Phase Transitions and Conformal Invariance* (Springer, Heidelberg 1999).
20. U. Niederer, Helv. Phys. Acta **45**, 802 (1972).
21. C.R. Hagen, Phys. Rev. D **5**, 377 (1972).
22. L. O'Raifeartaigh and V.V. Sreedhar, Ann. of Phys. **293**, 215 (2001).
23. M. Henkel, J. Unterberger, Nucl. Phys B **660**, 407 (2003).
24. M. Pleimling, M. Henkel, Phys. Rev. Lett. **87**, 125702 (2001).
25. R. Cherniha, M. Henkel, math-ph/0402059; J. Math. Anal. Appl., in press.
26. H.K. Janssen, in G. Györgyi et al. (eds) *From Phase transitions to Chaos*, World Scientific (Singapore 1992), p. 68
27. M. Henkel, M. Pleimling, C. Godrèche, J.-M. Luck, Phys. Rev. Lett. **87**, 265701 (2001).
28. A. Barrat, Phys. Rev. E **57**, 3629 (1998).
29. D.B. Abraham and P.J. Upton, Phys. Rev. B **39**, 736 (1989).
30. F. Corberi, E. Lippiello, M. Zannetti, Phys. Rev. E **68**, 046131 (2003).
31. M. Henkel, M. Pleimling, Phys. Rev. E **68**, 065101(R) (2003).
32. G.F. Mazenko, Phys. Rev. E **69**, 016114 (2004).
33. P. Calabrese, A. Gambassi, Phys. Rev. E **67**, 036111 (2003).
34. A.J. Bray, B. Derrida, Phys. Rev. E **51**, R1633 (1995).
35. A. Picone, M. Henkel, cond-mat/0402196; Nucl. Phys. B **688**, 217 (2004).
36. D. Giulini, Ann. of Phys. **249**, 222 (1996).
37. G. Burdet, M. Perrin, P. Sorba, Comm. Math. Phys. **34**, 85 (1973).
38. M. Henkel, A. Picone, M. Pleimling, cond-mat/0404464.

# Phase Transitions and Segregation Phenomena in Vibrated Granular Systems

Christof A. Kruelle<sup>1</sup>, Sébastien Aumaître<sup>2</sup>, Andreas P. J. Breu<sup>1</sup>,  
Andreas Goetzendorfer<sup>1</sup>, Tobias Schnautz<sup>1</sup>, Rafał Grochowski<sup>3</sup>, and  
Peter Walzel<sup>3</sup>

<sup>1</sup> Experimentalphysik V, Universität Bayreuth  
95440 Bayreuth, Germany

<sup>2</sup> Laboratoire de Physique Statistique, ENS  
75231 Paris, France

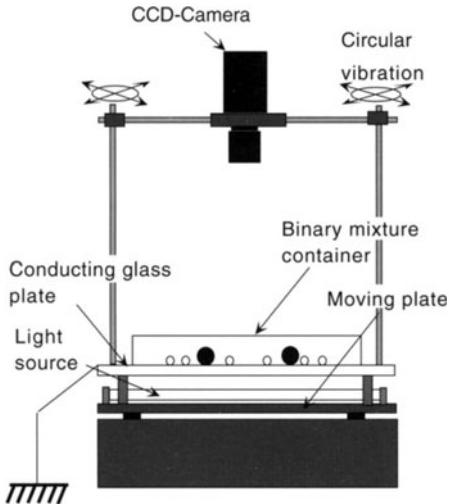
<sup>3</sup> Mechanische Verfahrenstechnik, Universität Dortmund  
44227 Dortmund, Germany

**Abstract.** A summary of results is presented from experiments in granular systems, which are excited by vertical and/or horizontal vibrations. The transitions between different dynamic states depend on internal properties of the granular system like the density of particles as well as external parameters of the driving shakers. Characteristic for granular systems are counterintuitive phenomena like the crystallization by increasing the vibration amplitude and thereby the energy input, or the rise of large particles in a sea of smaller ones (Brazil-nut effect). For horizontal shaking of a binary system the demixing of small and large particles is found to occur at the same critical particle density as the liquid-solid transition, which leads to the conclusion that both phenomena, segregation and phase transition, are closely related.

## 1 Introduction

An accumulation of macroscopic grains, set in motion by an external driving force, can show a surprising behavior. A peculiar phenomenon – called segregation – occurs as soon as heterogeneous particles are implied, in geophysical rock avalanches as well as in processed powders in the food or chemical industry [1,2]. This demixing of grains, which differ in size, density, or surface properties, has been intensively studied since the 1990s in laboratory experiments, e.g. in rotating drums [3] and under vertical [4] or, more recently, horizontal linear [5] excitation. Another phenomenon frequently encountered when handling granular material is the transition from a disordered phase to a more organized state, when the density of grains is increased beyond a critical value. This can be observed, for instance, in monodisperse particles under vertical vibration without gravity [6] or under horizontal translational excitation [7].

In this brief review we report on experimental investigations of both phase transitions and segregation phenomena in granular systems, which are agitated by three different vibration excitors. When the particles are externally



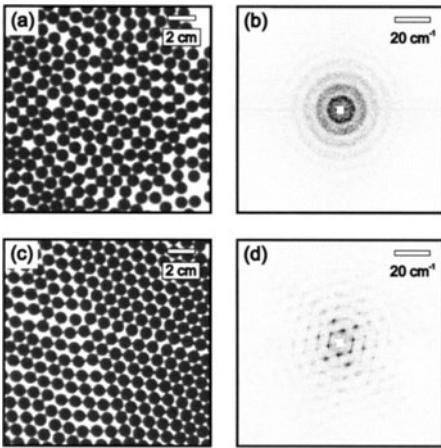
**Fig. 1.** Sketchy representation of the experimental device for the horizontal circular vibration of a monolayer of glass beads. The diameter of the container is 29 cm. The bottom glass plate is coated with an electrically conducting indium-tin oxide layer to suppress static electricity. From [8]

forced to perform stochastic movements a ‘granular temperature’ can be defined as mean kinetic energy in the center-of-mass system. For describing the physical properties of this ensemble one can ask – in analogy to thermodynamic phase transitions: (i) Are there critical temperatures at which the internal structure undergoes qualitative changes? (ii) Which order parameters characterize the dynamics of the transition? (iii) What are the consequences for granular mixtures?

## 2 Horizontal Circular Vibration

Here we present a model system, consisting of two species of glass beads with different size rolling in a horizontal shaker, where both phenomena, segregation and phase transition, are found to be closely related since they occur at about the same granular density [8].

A sketch of our experimental device is shown in Fig. 1. The particles in the dish are excited in a *horizontal circular vibration*, i.e. a circular movement of the entire platform due to the superposition of two sinusoidal vibrations in perpendicular directions. The frequency  $f$  of the table motion can be tuned from 0.5 to 2.0 Hz, with a preset amplitude  $A = \frac{n}{8} \cdot 2.54$  cm, where  $n = 2, \dots, 8$ . The granular system is composed of a various number,  $N$ , of spheres with diameters  $d = 0.4$  or  $1.0$  cm. To obtain a size-independent control parameter the filling fraction  $\mu$  is defined as the total cross-sectional area  $N \cdot \pi(d/2)^2$  of all spheres divided by the surface  $\pi(D/2)^2$  of the cell with diameter  $D = 29$  cm. During the motion, the positions of all particles are captured with a charge-coupled-device (CCD) camera fixed to the moving table.

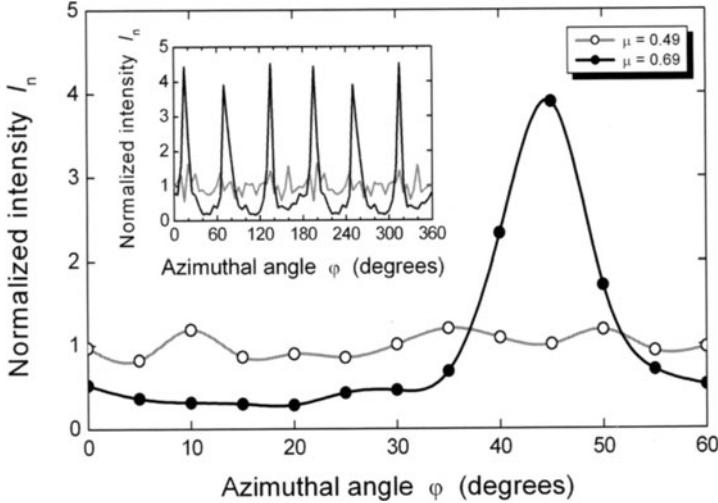


**Fig. 2.** (a) Image of a monolayer of 410 spheres (filling fraction  $\mu = 0.49$ ) rolling on a table shaken with an amplitude  $A = 2.22$  cm and a frequency  $f = 1.67$  Hz. (b) Contour plot of the corresponding 2D power spectrum. (c) Image of a monolayer of 580 spheres ( $\mu = 0.69$ ) under the same excitation, and (d) its 2D power spectrum. The snapshots are taken in a square of 15 cm at the center of the cell

During the experiments we noticed global changes in the dynamics of the system while crossing a critical threshold of the filling fraction  $\mu$ . In particular, the spheres became arranged in regular, triangular patches. To study this phenomenon, a monolayer of *monodisperse* glass spheres is placed inside the dish. After an early stage, during which the system loses all traces of the arbitrary initial configuration, an image is taken. To underline the developed structured state a 2D Fourier transformation (FFT) is performed. Results of these measurements are shown in Fig. 2, for low and high filling fractions of spheres. At low density, the configuration of grains does not show any structures. Its 2D power spectrum displays a continuous intensity distribution within a circle of radius  $k_d = 2\pi/d$ . In contrast, at high density, the small spheres arrange in a triangular lattice. In the 2D power spectrum six peaks appear at a wave number  $k_0 = 2\pi/\frac{\sqrt{3}}{2}d$ , evenly spaced by an angle of  $\pi/3$ .

The dynamics of the particles is obviously different in these two regimes. At low density each particle is free to follow its own trajectory until it collides with its neighbor, like in a fluid. On the other hand, at high density, the particles are forced into a collective motion inside a 2D crystal. The crossover between these two regimes is reminiscent of a liquid-solid transition.

To specify the transition point, a characteristic order parameter is extracted from the power spectra. The spectral intensity is integrated radially in an annulus  $0.98 k_0 < k < 1.02 k_0$  around the expected peaks. The resulting averaged intensity defines a function of the azimuthal angle  $\varphi$ . In the structured state this function is supposed to present six equidistant peaks, see inset of Fig. 3. To obtain a *single* characteristic number, which quantifies the order of the structured state, the angular space  $[0, 2\pi]$  is subdivided into six equal parts. The intensities in each interval are summed up yielding a singly



**Fig. 3.** Normalized intensity  $I_n$  of the power spectrum for  $\mu = 0.49$  ( $N = 410$ ) and  $\mu = 0.69$  ( $N = 580$ ), respectively. The absolute maxima  $I_{n,\max}$  are extracted to obtain the order parameter. The inset shows the full range of intensities within rings of same width around the centers of the 2D power spectra

peaked function  $I(\varphi)$  in the reduced angular range  $\varphi \in [0, \pi/3]$ . Dividing  $I(\varphi)$  by its arithmetic mean yields a normalized intensity

$$I_n(\varphi_i) = \frac{I(\varphi_i)}{\frac{1}{i_{\max}} \sum_{i=1}^{i_{\max}} I(\varphi_i)}, \quad \varphi_i = i \cdot \frac{\pi}{3i_{\max}}, \quad i_{\max} = 12,$$

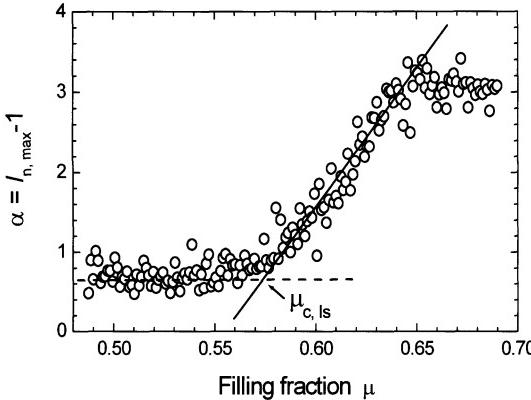
from which an order parameter

$$\alpha \equiv I_{n,\max} - 1$$

is derived. The open circles in Fig. 3 show that at a low density of spheres, where no coherent structure exists, no peak survives this analytic procedure ( $I_{n,\max} = 1.2$ ), whereas at a high density a pronounced peak remains (solid circles in Fig. 3). Its height  $I_{n,\max} = 4.0$  reflects the degree of symmetry established in the triangular lattice of the particles.

The values of  $\alpha$  versus the filling fraction  $\mu$  are shown in Fig. 4 for the driving conditions  $A = 2.22\text{ cm}$  and  $f = 1.67\text{ Hz}$ . A *liquid-solid like* transition can be observed around a critical filling fraction  $\mu_{c,ls}$ .

For the segregation experiments, a binary mixture of glass beads consisting of 19 large particles of diameter  $d_l = 1.0\text{ cm}$  embedded in a variable number of small spheres with diameter  $d_s = 0.4\text{ cm}$  are prepared. Initially the large spheres are placed on a regular triangular lattice with 3 cm spacing between the centers of two nearest neighbors. This distance  $L_{nn}(t)$ , averaged



**Fig. 4.** Evolution of the order parameter  $\alpha$  with the filling fraction  $\mu$ . The critical point  $\mu_{c,ls} = 0.57 \pm 0.01$  is found from the intersection of a linear fit of the data with  $1.5 < \alpha < 3.0$  with the mean level of the precritical data

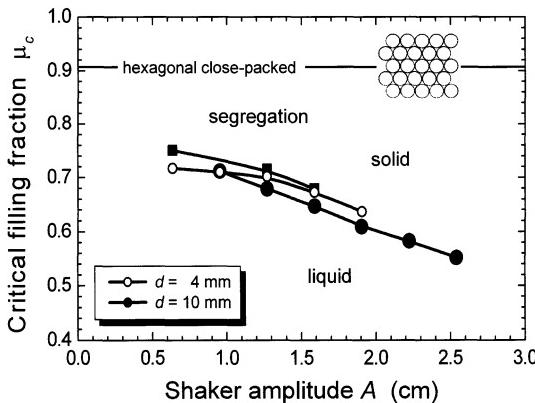
over all the 19 large spheres, is measured in real time [9]. For a clear characterization of the segregation the mean value of the distance between nearest neighbors in the asymptotic regime,  $L_\infty$ , is measured as a function of the filling fraction  $\mu$ . Here, a transition between a non-segregated state at low density and a segregated state at high density can be defined.

Since both transitions depend on the shaker amplitude we explored this dependence further. Fig. 5 shows the resulting phase diagram of the critical filling fractions  $\mu_{c,seg}$  and  $\mu_{c,ls}$ , respectively, vs. shaker amplitude  $A$ . These critical filling fractions seem to be independent on the driving frequency but decrease with increasing shaker amplitude. The regime where segregation occurs is always slightly above the liquid-solid transition line. This suggests that the granular phase transition is a precondition for segregation.

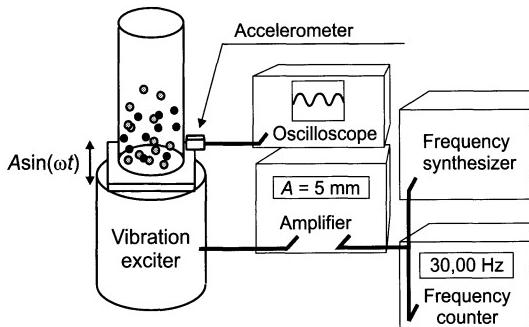
The transition is in the regime from 0.55 to 0.75 and depends on the amplitude of the driving shaker in a counterintuitive way: the liquid-solid transition line can be crossed by increasing the amplitude and thereby the energy input. A similar phenomenon was coined ‘freezing by heating’ in the context of driven mesoscopic systems [10].

We offer the following thoughts to elucidate the connection between the segregation in the binary mixture and the phase transition in the monodisperse layer of spheres. It seems that the granular material does not stay uniformly distributed but chooses the configuration which minimizes the energy input for a given number of particles and excitation parameters, in order to reduce the amount of energy to be dissipated. For  $\mu \leq \mu_c$  all particles move randomly with a uniform distribution, i.e., they prefer not to collide with the lateral wall. Indeed at low densities only few particle-wall collisions can be observed. We have measured that, in this case, large and small particles roll inside the dish in almost the same fashion. No segregation occurs. On the other hand, if the filling fraction exceeds  $\mu_c$ , a boundary-hitting regime

begins. This change can be seen and even heard in the experiment. In this regime the granules are continuously hit by the lateral wall. This process tends to decrease the extension of the monolayer which organizes itself in a more and more condensed state, leading to a triangular lattice, when the excitation is increased even further. Following the ideas of Edwards [11], this is also the reason why segregation occurs: beyond the critical filling fraction the collisions with the lateral wall become the dominant driving mechanism. The granular material is continuously compressed by the collisions with the lateral wall and therefore tends to increase its compactness by organizing itself in a triangular lattice. Since the larger particles disturb this ordering process the system pushes these intruders into the central region of the dish thereby reducing the number of holes and defects [12,13].



**Fig. 5.** Phase diagram of the critical filling fraction  $\mu_c$  vs. shaker amplitude  $A$  showing the regime  $\mu > \mu_{c,\text{seg}}$  (squares), where segregation occurs. The liquid-solid transition line  $\mu > \mu_{c,\text{ls}}$  (circles) is obtained from the normalized Fourier intensity data for two different particle sizes. Data partially taken from [8]



**Fig. 6.** Sketchy representation of the experimental device for the vertical vibration of a binary mixture of spheres. The diameter of the container is 9.4 cm

**Table 1.** Densities and diameters of the spherical particles used in the experiments. Data as used in [25]

Material	$\rho$ (g cm $^{-3}$ )	$d$ (mm)
Wood	0.8	8 12
Synthetic resin	1.1	
Polyurethane	1.2	10
Polypropylene	1.5	10
Glass	2.5	4 6 8 10
Aluminum	2.6	2 6
Steel	7.7	2 4 6
Bronze	8.5	10

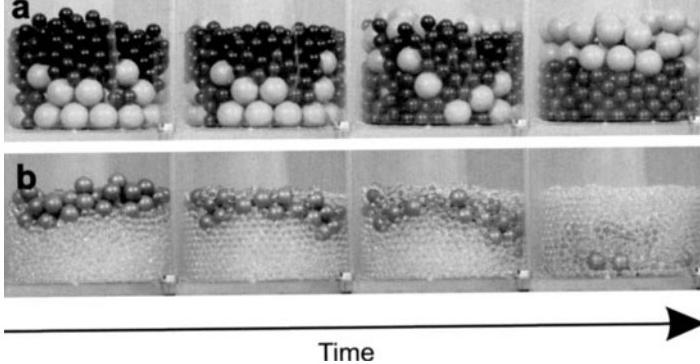
We conclude that the main mechanism for size segregation in our binary system is the compression force exerted by the lateral boundary and mediated by the developing lattice of the smaller spheres.

### 3 Vertical Vibration

As demonstrated in the previous section, granular media consisting of small and large particles tend to de-mix when shaken vigorously enough. The ‘Brazil-nut effect’ – if a particle mixture is shaken *vertically*, the larger particles will end up on top of the smaller ones, like the nuts in a muesli package – became the *Drosophila melanogaster* of granular media research [4,14,15,16,17]. Numerical simulations could validate the rise of larger particles [18,19]. Proposed mechanisms are, for example, convective motion of the smaller particles, which drag the larger ones to the top [15], or the filling of voids by the small particles, thereby lifting the larger ones [4,18,19]. Contrary to these common observations Shinbrot [20], however, noticed that a large particle, depending on its density, could also sink to the bottom of the container. Recent theoretical investigations [21,22,23] explained that both effects, the rise or descend of the larger particles, may occur. The latter case has become known as ‘reverse Brazil-nut effect’. The borderline between both effects has been predicted by Hong, Quinn, and Luding [21] in a simple relation between the size  $d_l/d_s$  and mass ratios  $m_l/m_s$  of the large and small particles.

One approach to describe an externally driven granular medium is to consider the individual particles as hard spheres. The driving could be – as in the case of the Brazil-nut effect – a vertical vibration of the container, which confines the granules. For strong driving, where all particles are in continuous motion, the ideas of the kinetic theory of gases can be applied. Then it is possible to define a granular temperature  $T$  in analogy to a gas, using the mean kinetic energy:

$$T \propto \frac{1}{N} \sum_i \frac{m_i}{2} (v_i - \langle v \rangle)^2, \text{ with } \langle v \rangle = \frac{1}{N} \sum_i v_i, \quad (1)$$



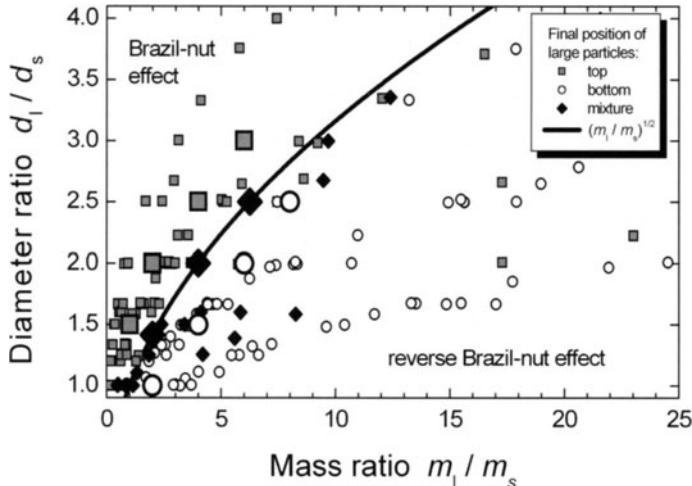
**Fig. 7.** Temporal evolution of the system. (a) Initially 8 mm glass beads on top of 15 mm polypropylene, which show the classical Brazil-nut effect. (b) 10 mm bronze spheres on 4 mm glass beads showing the reverse Brazil-nut effect. From [25]

where  $N$  is the total number of particles, and  $m_i$  and  $v_i$  are the mass and the velocity of the  $i$ -th particle, respectively. Hong [24] calculated, on the basis of this model, a critical granular temperature  $T_c = mgd\mu/\mu_0$ , below which a system of monodisperse spheres condensates. This critical temperature depends on the diameter  $d$  of the particles and the initial filling height  $\mu$ , measured in units of  $d$ .  $g$  denotes the gravitational acceleration and  $\mu_0$  is a constant, which depends on the spatial dimension and the underlying packing structure [24]. For a binary particle mixture different critical temperatures do exist, as pointed out by Hong *et al.* [21]. If a binary mixture of spheres is agitated by an external shaker, such that the granular temperature is in between the two critical values, one type of particle condensates while the other remains fluidized. It depends on the size and mass of the particles, which particle species will condensate and therefore sink to the bottom of the container. Following Rosato *et al.* [18], Hong *et al.* claimed that the cross-over condition is given when the ratio of the critical temperatures is equal to the volume ratio of the two particle species, which leads, in  $D$  dimensions, to the simple relation

$$\frac{d_l}{d_s} = \left( \frac{m_l}{m_s} \right)^{\frac{1}{D-1}}. \quad (2)$$

If, in  $3D$ , the diameter ratio is larger than the square root of the mass ratio, the particle mixture should show Brazil-nut effect and vice versa.

The prediction of formula (2) was put to the test [25] by preparing an instable layering, followed by a controlled shaking of the container. Our experimental device (see Fig. 6) operates at frequencies  $f$  between 0 and 100 Hz and normalized accelerations  $\Gamma = A(2\pi f)^2/g$  up to 40, where  $A$  is the shaking amplitude and  $g$  the gravitational acceleration. The acceleration  $\Gamma$  is measured with an accelerometer attached to the base plate of the Perspex

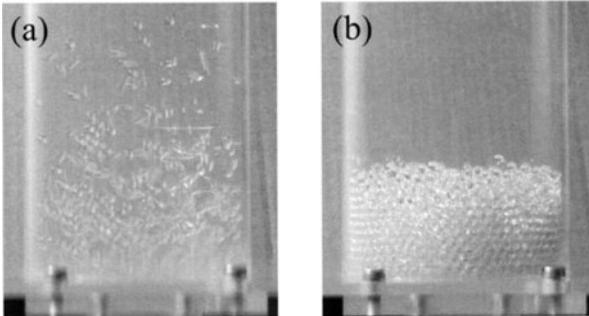


**Fig. 8.** Phase space for particle properties. The plot shows the regimes where reversed and classical Brazil-nut effects occur depending on the particle properties. Each small symbol represents one of 178 experiments (data taken from [25]). The solid line separating both areas is given by expression (2). The large symbols indicate the prediction of 3D molecular-dynamics simulations performed by Hong *et al.* [21] with up to 3600 particles

cylinder. The granular materials used were spherical particles with properties listed in Table 1.

In order to check whether a binary mixture would show Brazil-nut behavior or its reverse, we prepared a presumed unstable configuration, i.e., we put several layers of the type, which were predicted by (2) to rise during shaking on the bottom of the Perspex vessel. On top of these we placed one or more layers of the second type. If the prediction was correct, we would observe that the particles on top started to move through the layers of the other particle type, ending up at the bottom of the vessel (see Fig. 7). On the other hand, if the prediction turned out to be wrong the initial layering would be stable.

Most experiments have shown clearly either the Brazil-nut effect (Fig. 7(a)) or the reverse form (Fig. 7(b)). For some particle combinations a mixed state was stable. An overview of the results of the experimental tests as well as numerical simulations by Hong, Quinn, and Luding [21] together with their theoretical borderline are shown in Fig. 8. For 81% of the tested combinations (145 out of 178) the prediction of (2) was correct. The prediction failed when one particle type was made of aluminum or polyurethane. We assume that for these materials the condition of hard spheres, which is one of the main propositions in the theory was not met. We also noticed that, in the case of the reverse Brazil-nut effect, it is crucial to choose an appropriate filling height. It turned out that the effect is completely destroyed, if the initially lower layer is too large.



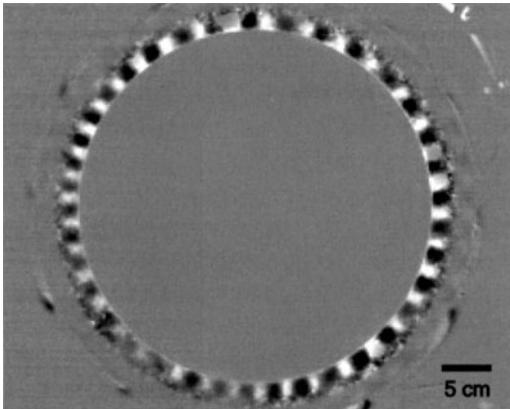
**Fig. 9.** Behavior of an ensemble of monodisperse glass beads (diameter = 4 mm) at fixed acceleration: (a) granular gas for subcritical filling, (b) crystalline block for supercritical filling

We currently focus on experimental verifications of the premise of this segregation mechanism via condensation. If, at a fixed external driving, a monodisperse set of glass beads exceeds a critical number of particles a phase transition can be observed from a fluidized ‘gas-like’ state to a condensed ‘crystalline’ state (see Fig. 9). Systematic studies of the dependence of the condensation temperature  $T_c$  on the internal parameters (number, size, and material density of the spheres) as well as the external conditions (amplitude and frequency of the shaker) are promising [26].

#### 4 Vertical Circular Vibration

The combined action of a horizontal *and* vertical vibration of the support has been utilized already for a long time for the controlled transport of bulk cargoes in a whole variety of industrial processes [27,28,29]. Since the transport phenomena on these so-called ‘vibratory conveyors’ involve the nonlinear interaction of many-particle systems with complex behavior leading to self-organized spatiotemporal patterns, the investigation of their dynamical properties has become a challenging subject to physicists, too.

In this final section we report on a conveyor system designed as a prototype apparatus specially developed to investigate the transport under principal oscillation modes, i.e. linear, elliptical, and circular for a long running time, without disturbing boundary conditions [31,30]. In our experiments, the transporting trough has the form of a horizontally oriented ring with radius  $R = 22.5\text{ cm}$ . The motion of the ring can be described by a trajectory performed on a cylindrical surface, consisting of a vertical oscillation  $z(t) = A \cos(\omega t)$  superposed with a torsional vibration  $\phi(t) = A/R \cos(\omega t + \varphi)$  around the symmetry axis of the apparatus, where  $\varphi$  is the fixed phase shift between the two oscillations. If, for example, the phase shift  $\varphi$  is chosen to be  $\pi/2$ , then each point on the trough traces a circular path in a vertical plane tangent



**Fig. 10.** Granular surface waves ( $\lambda = 3.8 \text{ cm}$ ) inside a vibratory conveyor with vertical circular motion of the annular trough oscillating with  $\Gamma = 3.0$  at a frequency  $f = 22.4 \text{ Hz}$

to the trough at that point. In short, the support agitates the granules via a *vertical circular vibration*.

The dimensionless acceleration  $\Gamma$  of the conveyor can be varied by changing the rotation frequency  $f$  of the driving units in the range  $0 < \Gamma < 7$ . The granulate used consists of  $\approx 520\,000$  glass beads with diameter  $d = 1.1 \text{ mm}$ , yielding a layer height of  $\approx 10 d$ .

Below a critical value  $\Gamma_c \approx 0.45$  the grains stay at rest, i.e., they follow the agitation of the tray without being transported. The onset of particle movements is restrained by frictional forces between grains and the substrate. For accelerations above this threshold the granular material sets in motion. Individual particles are unblocked and begin to move freely on top of each other. A net granular flow with constant velocity is observed.

If  $\Gamma$  exceeds 1, the vertical component of the circular acceleration will cause the grains to detach from the bulk followed by a flight on a ballistic parabola. This results in a much less dense-packed, ‘fluidized’, state with highly mobile constituents. In a certain driving range at around  $\Gamma = \pi$  a locking of the time-of-flight between successive bounces and the period of the circular vibration occurs. The initially flat bed becomes destabilized, and undulations of the granular surface occur [31]. In contrast to the surface waves observed earlier in purely vertically vibrating systems [32,33,34,35], here, the wave crests are not stationary but propagate along the trough. The relationship between the wave speed and the transport velocity of the particles is currently under investigation.

## 5 Conclusion

These three experiments indicate that the four major concepts describing the complex behavior of a vibrated granular system, namely phase transition, segregation, pattern formation, and transport are closely related and yield a rewarding field for future research.

## Acknowledgements

We would like to thank H. Elhor, A. García Sánchez, F. Landwehr, S.J. Linz, I. Rehberg, and M. Rouijaa, for valuable discussions. Support by Deutsche Forschungsgemeinschaft (DFG-Sonderprogramm ‘Verhalten granularer Medien’) and EU (TMR network ‘Patterns, Noise and Chaos’) is gratefully acknowledged.

## References

1. H.M. Jaeger, S.R. Nagel, and R.P. Behringer, Rev. Mod. Phys. **68**, 1259 (1996).
2. L.P. Kadanoff, Rev. Mod. Phys. **71**, 435 (1999).
3. K.M. Hill and J. Kakalios, Phys. Rev. E **49**, R3610 (1994).
4. J. Duran, J. Rajchenbach, and E. Clément, Phys. Rev. Lett. **70**, 2431 (1993).
5. P.M. Reis and T. Mullin, Phys. Rev. Lett. **89**, 244301 (2002).
6. E. Falcon, R. Wunenburger, P. Evesque, S. Fauve, C. Chabot, Y. Garrabos, and D. Beysens, Phys. Rev. Lett. **83**, 440 (1999).
7. G. Strassburger and I. Rehberg, Phys. Rev. E **62**, 2517 (2000).
8. S. Aumaître, T. Schnautz, C.A. Kruelle, and I. Rehberg, Phys. Rev. Lett. **90**, 114302 (2003).
9. S. Aumaître, C.A. Kruelle, and I. Rehberg, Phys. Rev. E **64**, 041305 (2001).
10. D. Helbing, I.J. Farkas, and T. Vicsek, Phys. Rev. Lett. **84**, 1240 (2000).
11. S.F. Edwards and D.V. Grinev, Phys. Rev. E **58**, 4758 (1998).
12. A.D. Dinsmore, A.G. Yodh, and D.J. Pine, Phys. Rev. E **52**, 4045 (1995).
13. J. Duran and R. Jullien, Phys. Rev. Lett. **80**, 3547 (1998).
14. J.C. Williams, Powder Tech. **15**, 245 (1976).
15. J.B. Knight, H.M. Jaeger, and S.R. Nagel, Phys. Rev. Lett. **70**, 3728 (1993).
16. M.E. Möbius, B.E. Lauderdale, S.R. Nagel, and H.M. Jaeger, Nature **414**, 270 (2001).
17. M. Medved, H.M. Jaeger, and S.R. Nagel, Phys. Rev. E **63**, 061302 (2001).
18. A. Rosato, K.J. Strandburg, F. Prinz, and R.H. Swendsen, Phys. Rev. Lett. **58**, 1038 (1987).
19. R. Jullien, P. Meakin, and A. Pavlovitch, Phys. Rev. Lett. **69**, 640 (1992).
20. T. Shinbrot and F.J. Muzzio, Phys. Rev. Lett. **81**, 4365 (1998).
21. D.C. Hong, P.V. Quinn, and S. Luding, Phys. Rev. Lett. **86**, 3423 (2001).
22. N. Shishodia and C.R. Wassgren, Phys. Rev. Lett. **87**, 084302 (2001).
23. J.A. Both and D.C. Hong, Phys. Rev. Lett. **88**, 124301 (2002).

24. D.C. Hong, *Physica A* **271**, 192 (1999).
25. A.P.J. Breu, H.-M. Ensner, C.A. Kruelle, and I. Rehberg, *Phys. Rev. Lett.* **90**, 014302 (2003).
26. A. García Sánchez, A.P.J. Breu, A. Goetzendorfer, C.A. Kruelle, and I. Rehberg, in preparation
27. G. Pajer, H. Kuhnt, and F. Kuhnt, *Fördertechnik – Stetigförderer* (VEB Verlag Technik, Berlin, 5th edn., 1988).
28. F.J.C. Rademacher and L. Ter Borg, *Eng. Res.* **60**, 261 (1994).
29. E.M. Sloot and N.P. Kruyt, *Powder Technol.* **87**, 203 (1996).
30. R. Grochowski, P. Walzel, M. Rouijaa, C. A. Kruelle, and I. Rehberg, *Appl. Phys. Lett.* **84**, 1019 (2004).
31. M. Rouijaa, C. Krüller, I. Rehberg, R. Grochowski und P. Walzel, *Chemie Ingenieur Technik* **76**, 62 (2004).
32. S. Douady, S. Fauve, and C. Laroche, *Europhys. Lett.* **8**, 621 (1989).
33. H.K. Pak and R.P. Behringer, *Phys. Rev. Lett.* **71**, 1832 (1993).
34. F. Melo, P. Umbanhowar, and H.L. Swinney, *Phys. Rev. Lett.* **72**, 172 (1994).
35. T. Metcalf, J.B. Knight, and H.M. Jaeger, *Physica A* **236**, 202 (1997).

# Ab-Initio Thermodynamics of Metal Alloys: From the Atomic to the Mesoscopic Scale

Stefan Müller

Universität Erlangen-Nürnberg, Lehrstuhl für Festkörperphysik  
Staudtstr. 7, 91058 Erlangen, Germany

**Abstract.** Density functional theory (DFT) based calculations allow us to study a number of metal alloy properties, as e.g. formation enthalpies, or electronic and elastic properties of intermetallic compounds. However, such so-called ab-initio calculations can only be applied as long as the alloy structure requires only small unit cells for the crystallographic description. It will be demonstrated how this limitation can be overcome without giving up the accuracy of DFT calculations by combining them with so-called Cluster Expansion (CE) methods and Monte-Carlo (MC) simulations. This concept gives access to systems containing more than a million of atoms and therefore, permits to treat alloy properties which possess a delicate temperature-dependence. As examples, we discuss mixing enthalpies, short-range order, and precipitation *without* any empirical parameters as input, but with an accuracy that allows for the *quantitative* prediction of experimental results.

## 1 Introduction

The increasing interest in metal alloys comes from their wide field of technical applications ranging from lightweight cars and aircraft turbine blades to protective coatings in corrosive environments and magnetic devices. Beside experimental studies which are sometimes connected with a tremendous financial effort, more and more computer based approaches attract attention by many research groups. The set of all such theoretical methods forms a new discipline known as “computational materials science”. Indeed, this expression includes a large number of different approaches reaching from continuum theory and semi-empirical accesses to so-called first-principles methods, mostly based on density functional theory [1,2]. Since the latter is strictly based on quantum-mechanical rules, it has a real predictive power, and –from a physical point of view– seems to be the most elegant solution to treat alloy problems. One should, however, keep in mind that we are already confronted with a many-body problem on a atomic scale and now, have to find a way, how such calculations may be used to describe properties which are directly connected to the phase stability of an alloy, as e.g. short-range ordering. In principle, this demands to solve the following three problems:

(i) *The problem of bridging length scales:* Many material properties take place on a mesoscopic or even macroscopic scale, i.e. model systems consisting of millions of atoms are necessary –a number far beyond the capacity of today’s computers.

(ii) *The problem of controlling huge configuration spaces:* The validity of a predicted ground-state diagram sensitively depends on the number of considered structures. Compared to the  $2^N$  configurations of a binary alloy system containing  $N$  atoms, DFT calculations are restricted to a tiny part of this configuration space. So, normally a set of “intuitive candidates” is chosen and that with the minimum energy is postulated as ground-state. Naturally, this procedure fails to allow for surprises.

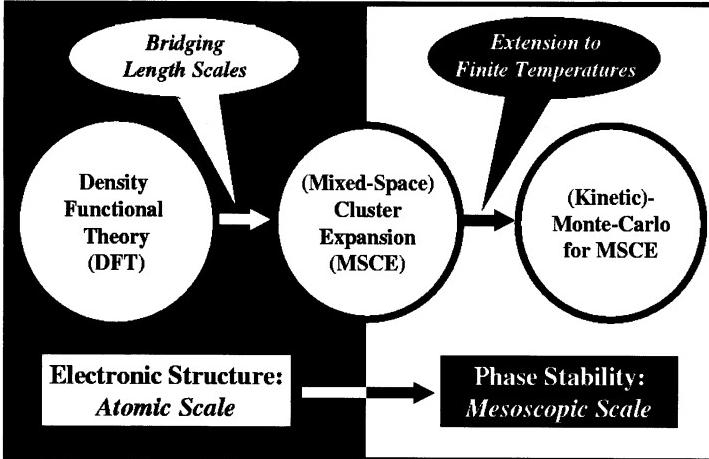
(iii) *The problem of treating configurational entropy:* Ordering phenomena of alloys are temperature-dependent. This dependence cannot be treated by methods which only search the positional space. Instead, the configurational space must be taken into account allowing for exchange processes between individual atoms.

It will be shown, how these problems can be overcome by combining DFT-based methods with concepts from statistical physics. Our focus will be on binary metal alloy systems. The structure of the present work is as follows: First the theoretical methods will be introduced (Section 2). Then, the concept will be applied to two different alloy properties, namely precipitation (Section 3) and short-range order (Section 4). Finally, the limits and future possibilities will be briefly discussed in form of a summary (Section 5).

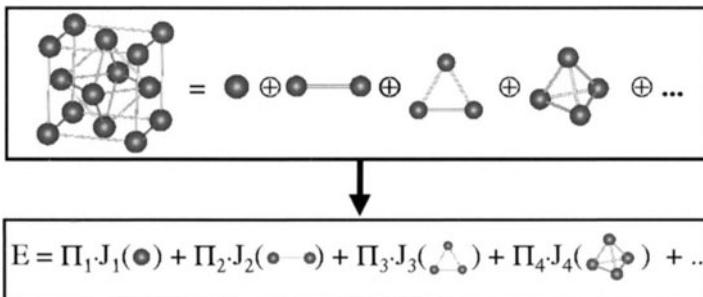
## 2 Methods

The concept of our theoretical approach is shown in Fig. 1. The starting point builds DFT which is based on the Hohenberg-Kohn-theorem [1] stating that the energy of a system of interacting electrons in an external potential depends only on the ground state electronic density. In our case, namely the investigation of solid structures, the external potential is the Coulomb potential caused by the nuclei in a solid. Since there exists a number of excellent review articles [3,4] and books [5,6,7] about DFT, no further details are given here. In a second step, the accuracy of DFT is extended to huge configuration spaces by combining DFT with concepts from statistical mechanics. The basic idea by Sanchez, Ducastelle and Gratias [8] is called “Cluster Expansion”(CE), and sketched in Fig. 2: For a given underlying lattice, the crystal structure is divided into characteristic figures such as pairs, triangles, etc. Then, the energy of *any* configuration  $\sigma$  on this lattice can be uniquely written [8] as linear combination of the characteristic energies  $J$  of each individual figure. In practice, the only error we make is that the sum must be truncated at some point. The  $\Pi$ 's in Fig. 1 are structure-dependent factors.

In practice, the construction of the so-called effective interactions  $J$  is done by an inversion method [11]. This demands the calculation of the energies for a set (typically 15-50) of *geometrically fully relaxed* structures via DFT which are then used to find that set of interactions,  $\{J\}$ , which minimize the deviation between energies resulting from CE and DFT for all structures  $\sigma$ . This procedure has to result in effective cluster interactions  $J$  which



**Fig. 1.** The combination of DFT plus MSCE [9] plus MC [10] allows for a description of phase stability of metal alloys on a mesoscopic scale without any empirical parameters



**Fig. 2.** The concept of cluster expansions: The crystal is separated in characteristic figures (here, shown for the fcc-lattice). The energy of any configuration can then be written as linear combination of the characteristic energies  $J_f$  of the figures

on the average only causes an additional prediction error of 1-2 meV/atom compared to the direct DFT energy. For this, the set of DFT input structures must be varied until the effective interactions becomes stable, i.e. will not longer change, if an input structure is removed or added to the set of structures (for details, see e.g. Ref. [12]). Moreover, as first shown by Laks et al. [13], any CE in real space *fails* to predict the energy of long periodic coherent superlattices – a prerequisite for studying precipitation on a mesoscopic scale from first-principles. In principle, the problem can be solved by transforming a group of interactions to the reciprocal space which is easiest to do for the pair interactions and considering the long-periodic limit separately [13]. Finally, this leads to the Hamiltonian of the “Mixed-Space Cluster Expansion (MSCE)” [9,13]. Here, any configuration  $\sigma$  is defined by

specifying the occupations of each of the  $N$  lattice sites by an  $A$  atom (spin-index  $\hat{S}_i = -1$ ) or a  $B$  atom ( $\hat{S}_i = +1$ ). The formation enthalpy of any configuration  $\sigma$  at its atomically relaxed state is then given by

$$\Delta H_{CE}(\sigma) = \sum_{\mathbf{k}} J_{pair}(\mathbf{k}) |S(\mathbf{k}, \sigma)|^2 + \sum_f^{MB} D_f J_f \bar{H}_f(\sigma) + \\ + \frac{1}{4x-1} \sum_{\mathbf{k}} \Delta E_{CS}^{eq}(\hat{k}, x) |S(\mathbf{k}, \sigma)|^2. \quad (1)$$

The first two terms represent the *chemical energy*,  $E_{chem}$  (often referred to as *interfacial energy*). Here, the first sum describes all possible pair figures.  $J_{pair}(\mathbf{k})$  is the lattice Fourier transform of the real space pair interactions, and  $S(\mathbf{k}, \sigma)$  are structure factors. The second sum describes many-body figures, such as triangles, tetrahedra, etc.  $J_f$  is the real-space effective many-body interaction of figure  $f$ ,  $D_f$  stands for the number of equivalent clusters per lattice site and  $\bar{H}_f(\sigma)$  are spin products. The third term, the *constituent strain*,  $E_{CS}$  (often referred to as *coherency strain*), describes the strain energy necessary to maintain coherency between bulk element A and B along an interface with orientation  $\hat{k}$ . It can be calculated by deforming the bulk elements from their equilibrium lattice constants  $a_A$  and  $a_B$  to a common lattice constant  $a$  perpendicular to  $\hat{k}$ . The constituent strain is a function of composition  $x$  and direction  $\hat{k}$  only, but does not include information about the strength of chemical interactions between A and B atoms.

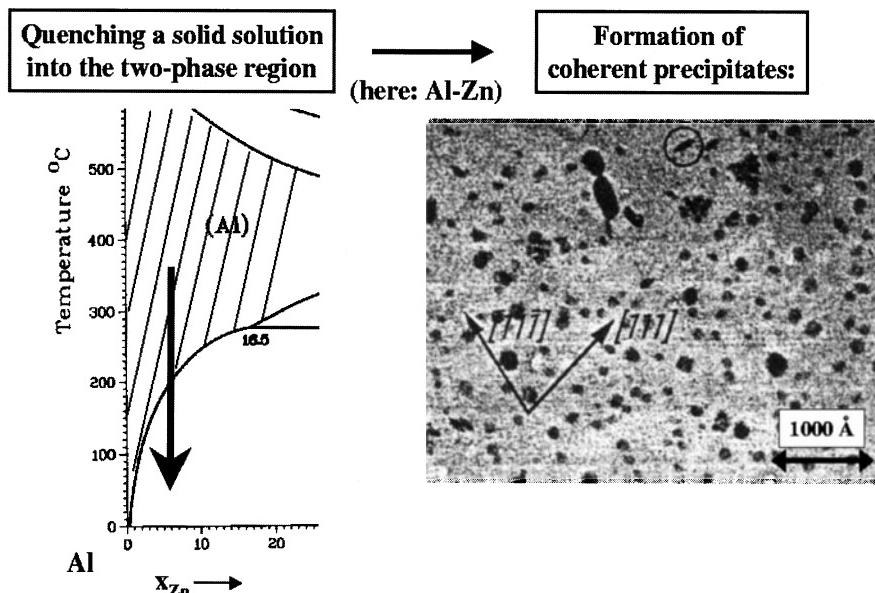
Although the CE ansatz allows to treat physical problems ranging from few up to a million atoms without giving up the accuracy of modern DFT calculations, the concept still needs to be extended to finite temperatures. This can be done by using the MSCE Hamiltonian, Eq. (1), in Monte-Carlo (MC) simulations (Fig. 1). We even can extend the field of applications to non-equilibrium processes by switching from thermodynamic MC to kinetic MC [14]. For this, activation barriers for exchange processes between neighbor atoms have to be considered, and the number of possible exchange processes between atoms must be restricted to small atomic distances. For phase-separating processes, as the formation and aging of precipitates which will be studied next, a MC code was created which is strongly related to the concept of the so-called “residence time algorithm” [15]. Here, atoms are forced to perform exchange processes and the time correlated to this process is calculated afterwards [10]. This algorithm is especially efficient for simulations of long aging times (many hours or days).

### 3 Precipitation

Solid state decomposition reactions like phase separation of an alloy into its constituents,  $A_{1-x}B_x \rightleftharpoons (1-x)A + xB$ , create so-called precipitates which define an important part of the microstructure of many alloy systems. The

early stage of these reactions typically involves the formation of *coherent precipitates* that adopt the crystallographic lattice of the alloy from which they emerge [16]. Coherent precipitates have practical relevance, as they impede dislocation motion, and thus lead to “precipitation-hardening” in many alloys [16,17,18,19]. So, their size vs. shape distribution as function of temperature and aging time is of special interest. Despite their importance, precipitate microstructures were thus far not amenable to first-principles theories, since their description requires “unit cells” containing  $10^3 - 10^6$  atoms or more. In experiment, precipitates are formed by quenching a solid solution deep into the two phase-region of the phase diagram, followed by sample aging as shown in Fig. 3 for fcc-based Al-rich Al-Zn alloys. The formed coherent precipitates can then be observed via transmission electron microscopy (TEM)[20]. The black spots are coherent fcc-Zn precipitates.

Following the classical textbook of Khachaturyan [18] the shape of a precipitate is controlled by two competing energies: While the interfacial or chemical energy leads to a compact shape, the strain energy leads to a flattening along the elastically soft direction of the precipitate. Our MSCE Hamiltonian easily allows for a separation into these two characteristic energy parts: As already mentioned in Section 2, the first two terms of Eq. (1) includes information about strength and importance of the individual pair and multi-body interactions and therefore, represents the chemical energy of the system. The last term, however, reflects the elastic properties of the alloy. We used

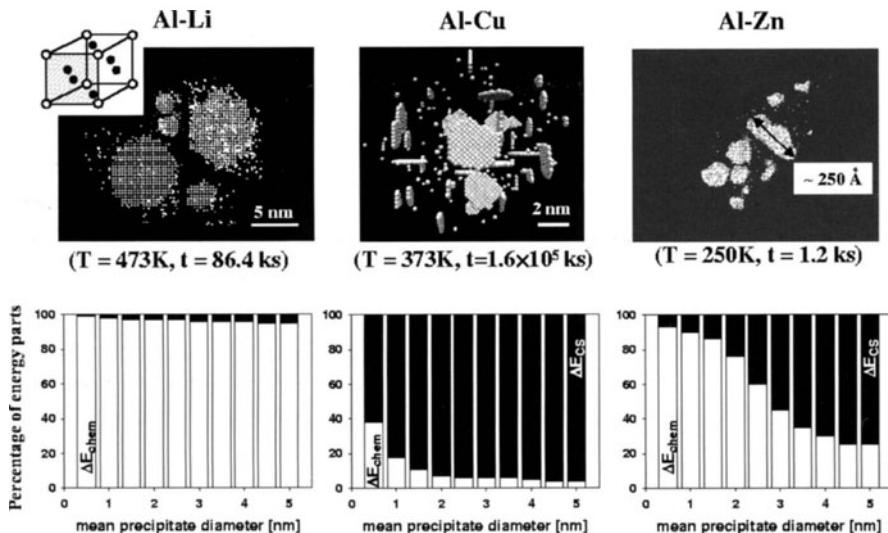


**Fig. 3.** Quenching an Al-rich Al-Zn solid-solution into the two phase region of the phase diagram (left), followed by sample aging, leads to the formation of Zn-precipitates which can then be observed via TEM [20] in form of black spots (right)

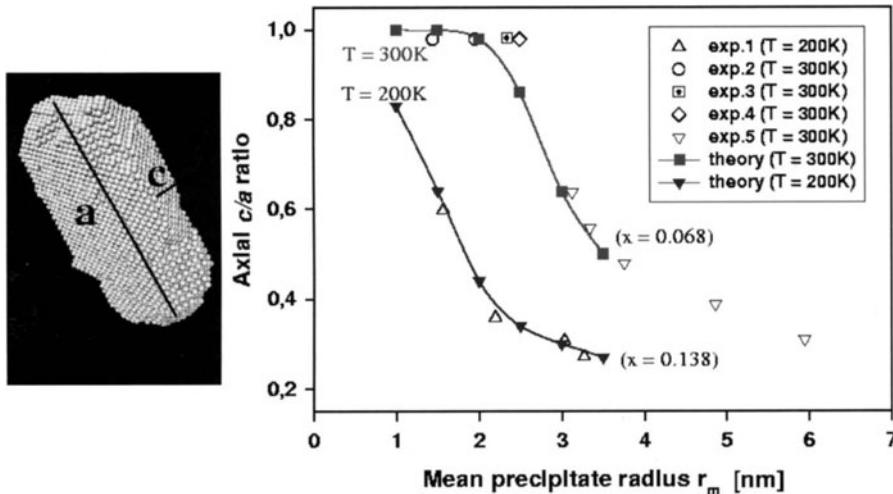
this simple physical picture to analyze the ratio between chemical energy  $E_{chem}$  and strain energy  $E_{CS}$ , as function of the precipitate size for three different Al-rich fcc based binary alloy systems, namely Al-Li, Al-Cu, and Al-Zn. The precipitate distributions are shown in Fig. 4, whereby only the Li, Cu, Zn atoms are shown, respectively. For the MC simulations model systems with up to one million atoms were necessary in order to receive sufficient statistics.

In the case of Al-Li, precipitates never flatten, but always possess a spherical shape. This becomes clear by analyzing  $E_{chem}/E_{CS}$  for different precipitate sizes as shown below the precipitate distribution images in Fig. 4: In the case of Al-Li, for all precipitate sizes the chemical energy  $E_{chem}$  (white bars) clearly dominates over the strain energy  $E_{CS}$  (black bars). The Li atoms seem to form a simple cubic lattice. This is due to the fact that the precipitates themselves show a  $\text{Al}_3\text{Li}$  stoichiometry forming the  $L1_2$  structure sketched on the top left corner of Fig. 4: While all corners of the unit cell are occupied by Li atoms, all faces are occupied by Al atoms. Since only the Li atoms are displayed in the real-space image, they form a simple cubic lattice. This observation also makes clear that there are practically no anti-phase boundaries within the  $\text{Al}_3\text{Li}$  precipitates which would demand the occupation of Al sites by Li atoms. Analyzing  $E_{chem}/E_{CS}$  for Al-Cu, we find the opposite is true than for Al-Li: The platelet-stabilizing strain energy is now the dominating part. So, precipitates flatten along the elastically soft direction of Cu being the [100] direction [21]. In the case of Al-Zn, the situation is more complex: Now, we see a strong size-dependence of the precipitate shape: Zn precipitates up to about 2.5 nm are more compact, i.e. chemically dominated, while larger precipitates becomes more and more ellipsoidal (strain dominated) [22].

Since the most interesting size-shape relation is found for Al-Zn, we will now focus on this system. One remarkable feature of its coherent precipitates is the fact that their short axis is always along the [111] (and symmetric equivalent) directions. Indeed, at a first glance, this appears a bit as a surprise, because most fcc elements are elastically soft along the [100] direction, and consequently hard along [111]. However, it should not be forgotten that we have to inquire about the stability and elastic properties of an unusual phase: fcc-Zn: While Zn is stable in the hcp structure, fcc-Zn shows an instability when deformed rhombohedrally along [111] [23]. As a consequent, fcc-Zn precipitates flatten along this direction [22]. This feature allows the definition of a  $c/a$  ratio and therefore, a quantitative measure for the description of the precipitate shape as used in many experimental studies and schematically shown on the left side of Fig. 5: While  $a$  represents the long axis of the ellipsoid (perpendicular to [111]),  $c$  is its thickness (parallel to [111]). The size is given by the radius of the associated sphere having the same volume as the corresponding precipitate. Figure 5 compares the experimental size-shape relation for two different aging times and concentrations [24,25,26,27,28] with those predicted from our method. For both temperatures, the agreement is



**Fig. 4.** Size-shape distribution of precipitates in Al-rich fcc based Al-Li, Al-Cu, and Al-Zn alloys (no Al atoms are shown) and their corresponding percentage of strain and chemical energy as function of the precipitate size. Precipitates of Al-Li form the  $L1_2$  structure as shown in the top left corner



**Fig. 5.** Shape ( $c/a$ ) vs. size relation of Zn precipitates for two different temperatures. The lines denote the results from our calculations [29], the open points are taken from different experimental studies (exp.1-5 correspond to Ref. [24,25,26,27,28])

excellent, i.e. our ansatz allows for a *quantitative* prediction of the size versus shape versus temperature relation of coherent precipitates [29].

## 4 Short-Range Order and Mixing Enthalpies

In general, all solid solutions of binary metal alloys show temperature-dependent short-range order (SRO). The MSCE Hamiltonian permit us to calculate SRO quantitatively by use of the Warren-Cowley SRO parameters given by

$$\alpha_{lmn}(x) = 1 - \frac{P_{lmn}^{A(B)}}{x} \quad (2)$$

where  $P_{lmn}^{A(B)}$  is the conditional probability that given an  $A$  atom at the origin, there is a  $B$  atom at  $(lmn)$ . For comparison with experimental data, the so-called “shells”  $lmn$  are introduced which are defined by the distance between  $A$  and  $B$  atoms in terms of half lattice parameters,  $(l\frac{a}{2}, m\frac{a}{2}, n\frac{a}{2})$ , e.g. for an fcc-lattice the nearest-neighbor distance would be described by the shell  $(110)$ , the second neighbor distance by  $(200)$  and so on. The sign of  $\alpha$  indicates whether atoms in a given shell prefer to order ( $\alpha < 0$ ) or cluster ( $\alpha > 0$ ). The SRO parameter may be written in terms of the cluster expansion pair correlations as [30]

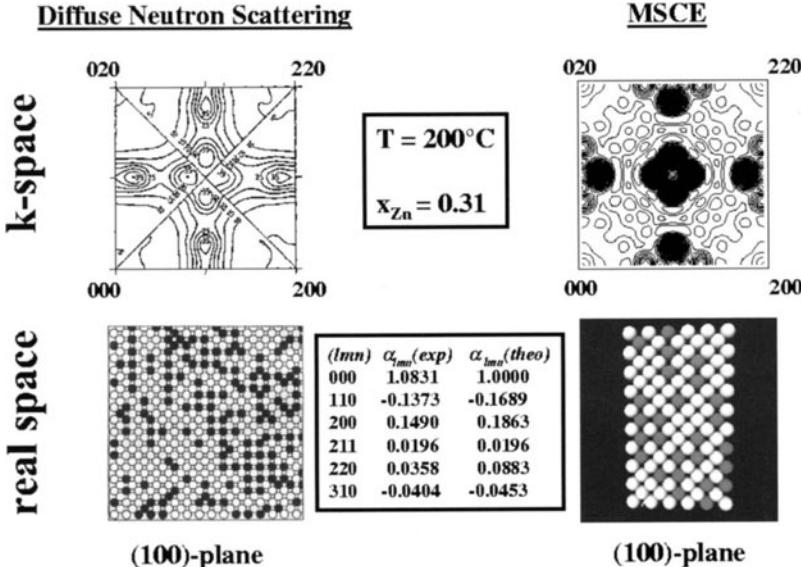
$$\alpha_{lmn}(x) = \frac{\langle \bar{\Pi}_{lmn} \rangle - q^2}{1 - q^2} \quad (3)$$

where  $q = 2x - 1$  and  $\langle \bar{\Pi}_{lmn} \rangle$  are the mentioned structure-dependent factors (see Section 2). In diffraction experiments the diffuse scattering due to SRO is proportional to the lattice Fourier transform of  $\alpha_{lmn}(x)$  [31,32]

$$\alpha(x, \mathbf{k}) = \sum_{lmn}^{n_R} \alpha_{lmn}(x) e^{i \cdot \mathbf{k} \cdot \mathbf{R}_{lmn}} \quad (4)$$

where  $n_R$  stands for the number of real space shells used in the transform. Equation (3) together with (4) opens the possibility to compare both, experimental and theoretically predicted diffuse diffraction patterns (reciprocal space) and SRO-parameters (real space).

As example, Fig. 6 compares experimental [33] and theoretical SRO patterns for  $\alpha$ -brass [34]. The patterns correspond to a sample with a Zn-concentration of 31.1% at  $T = 200^\circ\text{C}$ . Both patterns show SRO intensities around  $\mathbf{k} = \langle 1\frac{1}{4}0 \rangle$ . This reciprocal space vector is characteristic for the so-called  $D0_{23}$  structure which was earlier predicted as low-temperature ground state in  $\alpha$ -brass [34]. The structure belongs to the class of so-called long-periodic superlattices which are formed from the  $L1_2$  structure (which is sketched in Fig. 4, top left corner) by anti-phase boundaries along the  $[001]$  direction. Their fundamental reciprocal space vector is  $\mathbf{k} = \langle 1\frac{1}{2M}0 \rangle$  with  $M$

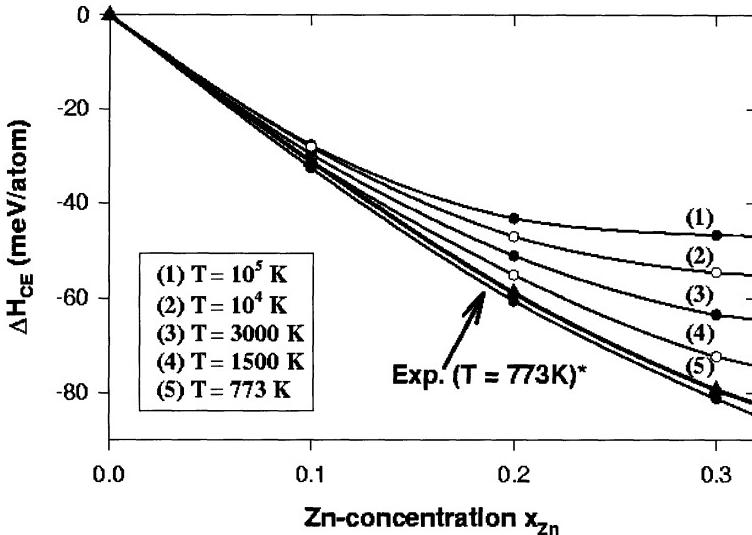


**Fig. 6.** SRO-pattern for a Cu-Zn solid solution ( $T = 200^\circ\text{C}$ ,  $x_{\text{Zn}} = 0.31$ ) resulting from neutron diffraction experiments [33] and from our calculation [34]. In both cases, SRO peaks around  $\langle 1\frac{1}{4}0 \rangle$  are visible favoring  $D0_{23}$  as low-temperature ground state. Lower part: Corresponding SRO parameters and real space images Chains of Zn-atoms along [001] can be seen

being the modulation wavelength. So, we see the highest intensity for the LPS structure with  $M = 2$  being the  $D0_{23}$  structure.

The SRO parameters can now be determined by a Fourier transformation of these SRO patterns. In experiment, they were used to construct a real space image of the alloy by using them as input for an inverse MC approach in order to obtain characteristic interactions [33]. Figure 6 compares the real space structure deduced from experiment [33] and from our calculation. In both cases, chains of Zn atoms are visible along [001], indicating that short-range order is essential for a quantitative correct description of the physical properties of the disordered solid-solution of  $\alpha$ -brass.

For a quantitative comparison of the computed SRO with experiment, Fig. 6 also gives the Warren-Cowley SRO parameter  $\alpha_{lmn}$  (Eq. (3)) for the first 5 shells. Considering the fact that the experimental error of  $\alpha_{000}$  amounts to be as large as 8% (since  $\alpha_{000} = 1.000$  by definition), the predicted and experimentally determined values agree very well. We see that  $\alpha_{110}$  is negative, indicating that Zn atoms prefer Cu atoms as nearest neighbors. Since SRO plays an important role in  $\alpha$ -brass, it should also influence its energetics, i.e. the mixing enthalpy of the system should possess a temperature-dependence. To clarify this, Fig. 7 compares calculated mixing enthalpies,  $\Delta H_{\text{mix}}(x, T)$ , for different temperatures with experimental data [35]. We



**Fig. 7.** Calculated mixing enthalpies of  $\alpha$ -brass for different temperatures [34] and comparison with experimental data [35] (bold line)

start with the random alloy which is realized by an unphysically high temperature ( $T = 10^5$  K), and go down to temperatures where SRO sets in. Comparing the energy curves for the random and the disordered alloy, we see that the calculation neglecting SRO leads to much higher mixing enthalpies. For higher Zn concentrations a good agreement between experiment and calculated mixing enthalpies can only be reached, if SRO is taken into account.

## 5 Summary

The applied combination of quantum mechanics (DFT) and statistical physics (MSCE, MC) certainly represents one of the most successful approaches to study binary alloy properties *without any empirical parameters*. At the moment, the limitation of the presented access is given by the underlying lattice which does, for example, not permit to study melting processes. Regarding ordering phenomena in the solid phase, the presented results allow for a quantitative prediction of experimental data. We have focused on two important alloy properties namely precipitation and short-range order (SRO). For the latter, it could be seen that a *quantitative* experiment-theory agreement can only be achieved, if the influence of SRO on the energetics of the disordered alloy is taken into account. This was demonstrated for  $\alpha$ -brass, where an excellent agreement between experimental and calculated mixing enthalpies could be reached. As an important example for a decomposition reaction, precipitation in different fcc-based Al-rich solid solution were studied. Here, we profit by the fact that the approach applied allows to study configurations

consisting of millions of atoms. In the case of Al-Zn, it was demonstrated that our prediction even goes through a quantitative comparison of the size-shape-temperature relation with experimental results.

The presented method is by no means restricted to the two characteristic ordering phenomena, but can also be applied to e.g. surface problems as adsorbate systems [36] or surface segregation [37]. Its application potential is up to now not yet clear. So, for example, after some further developments, it should be possible to describe properties as e.g. nucleation or island growth at surfaces.

### Acknowledgements

Special thanks go to Chris Wolverton, Alex Zunger, Lin-Wang Wang, Walter Wolf, Raimund Podloucky, and Long-Qing Chen for helpful discussions. Part of this work was supported by the US Department of Energy.

### References

1. P. Hohenberg and W. Kohn, Phys. Rev. **136**, 864B (1964).
2. W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 A (1965).
3. M. C. Payne, M. P. Teter, D. C. Allen, T. A. Arias, J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1064 (1992).
4. R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).
5. R. M. Dreizler, E. K. U. Gross, *Density Functional Theory* (Springer, Berlin, 1990).
6. R. M. Dreizler and J. da Providencia, *Density Functional Theory* (Plenum-Press, New York, 1985).
7. E. S. Krachko and E. V. Ludena, *Energy Density Functional Theory of Many electron Systems* (Kluwer Academic, Boston, 1990).
8. J. M. Sanchez, F. Ducastelle, D. Gratias, Physica A **128**, 334 (1984).
9. A. Zunger, NATO ASI on *Statics and Dynamics of Alloy Phase Transformations*, eds. P. E. A. Turchi and A. Gonis, Plenum Press, New York, 1994, p. 361.
10. S. Müller, L.-W. Wang, and A. Zunger, Model. Sim. Mat. Sci. Eng. **10**, 131 (2002).
11. J. W. D. Conolly and A. R. Williams, Phys. Rev. B **27**, 5169 (1983).
12. S. Müller, J. Phys.: Condens. Matt. **15**, R1429-R1500 (2003).
13. D. B. Laks, L. G. Ferreira, S. Froyen and A. Zunger, Phys. Rev. B **46**, 12587 (1992).
14. A. F. Voter, Phys. Rev. B **34**, 6819 (1986).
15. A. B. Börz, M. H. Kalos, and J. L. Lebowitz, J. Comp. Phys. **17**, 10 (1975).
16. A. Guinier, Solid State Physics **9**, 293 (1959).
17. J. B. Cohen, Solid State Physics **39**, 131 (1986).
18. A. G. Khachaturyan, *Theory of Structural Transformations in Solids* (John Wiley, New York, 1983).
19. H. W. Zandberger, S. J. Andersen, and J. Jansen, Science **277**, 1221 (1997).
20. R. Ramlau and H. Löffler, Phys. Stat. Sol. (a) **79**, 141 (1983).

21. C. Wolverton, Phil. Mag. Lett. **79**, 683 (1999).
22. S. Müller, C. Wolverton, L.-W. Wang, and A. Zunger, Acta Mater. **48**, 4007 (2000).
23. S. Müller, L.-W. Wang, and A. Zunger, C. Wolverton, Phys. Rev. B **60**, 16448 (1999).
24. G. Laslaz, P. Guyot, Acta Metall. **25**, 277 (1977).
25. E. Bubeck, V. Gerold, G. Kostorz, Cryst. Res. Technol. **20**, 97 (1985).
26. M. Fumeron, J.P. Guillot, A.P. Dauger, and J. Caisso, Scripta Metall. **14**, 189 (1980).
27. J. Deguercy, M. F. Denanot, M. Fumeron, J. P. Guillot, et J. Caisso, Acta Mater. **30**, 1921 (1982).
28. V. Gerold, W. Siebke, and G. Tempus, Phys. Stat. Sol. (a) **104**, 213 (1987).
29. S. Müller, C. Wolverton, L.-W. Wang, and A. Zunger, Europhys. Lett. **55**, 33 (2001).
30. V. Ozoliņš, C. Wolverton and A. Zunger, Phys. Rev. B **57**, 4816 (1998).
31. M. A. Krivoglaz, *X-Ray and Neutron Diffraction in Nonideal Crystals* (Springer, Berlin, 1996).
32. C. J. Sparks and B. Borie, in *Local Arrangements Studied by X-Ray Diffraction*, eds. J. B. Cohen and J. E. Hilliard (Gordon and Breach, New York, Met. Soc. Conf. **36**, 5 (1966)); B. Borie and C. J. Sparks, Acta Cryst. **17**, 827 (1964).
33. L. Reinhard, B. Schönfeld, and G. Kostorz, W. Bührer, Phys. Rev. B **41**, 1727 (1990).
34. S. Müller and A. Zunger, Phys. Rev. B **63**, 094294 (2001).
35. R. Hultgren, P. D. Desai, D. T. Hawkins, M. Gleiser, K. K. Kelley, *Selected Values of the Thermodynamic Properties of Binary Alloys* (American Society for Metals, Ohio, 1973).
36. R. Drautz, R. Singer, and M. Fähnle, Phys. Rev. B **67**, 035418 (2003).
37. O. Wieckhorst, S. Müller, L. Hammer, and K. Heinz, Phys. Rev. Lett., **92**, 195503 (2004).

# Theory of Collective Dynamics in Multi-Agent Complex Adaptive Systems

Neil F. Johnson<sup>1</sup>, Sehyo C. Choe<sup>1</sup>, Sean Gourley<sup>1</sup>,  
Timothy Jarrett<sup>1</sup>, and Pak Ming Hui<sup>2</sup>

<sup>1</sup> Physics Department, Clarendon Laboratory, Oxford University, UK

[n.johnson@physics.ox.ac.uk](mailto:n.johnson@physics.ox.ac.uk)

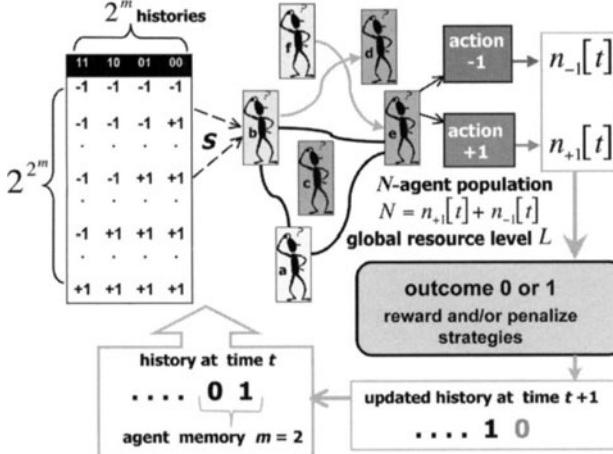
<sup>2</sup> Department of Physics, Chinese University of Hong Kong, Hong Kong

**Abstract.** We discuss a crowd-based theory for describing the collective behavior in Complex Adaptive Systems comprising multi-agent populations competing for a limited resource. These systems – whose binary versions we refer to as B-A-R (Binary Agent Resource) systems – have a dynamical evolution which is determined by the aggregate action of the heterogeneous, adaptive agent population. Accounting for the strong correlations between agents' strategies, yields an accurate analytic description of the system's dynamics.

## 1 Introduction

Complex Adaptive Systems – together with their dynamical behavior known as Complexity – are thought to pervade much of the natural, informational, sociological, and economic world [1,2]. Complex Adaptive Systems are probably best defined in terms of a list of common features which distinguish them from ‘simple’ systems, and from systems which are just ‘complicated’ as opposed to being complex. A list of Complex Adaptive System ‘stylized facts’ should include: feedback and adaptation at the macroscopic and/or microscopic level, many (but not too many) interacting parts, non-stationarity, evolution, coupling with the environment, and observed dynamics which depend upon the particular realization of the system.

Casti has argued that [1] ‘.... a decent mathematical formalism to describe and analyze the [so-called] El Farol Problem would go a long way toward the creation of a viable theory of complex, adaptive systems’. The rationale behind this statement is that the El Farol Problem, which was originally proposed by Brian Arthur [3] to demonstrate the essence of Complexity in financial markets involving many interacting agents, incorporates the key features of a Complex Adaptive System in an everyday setting. Very briefly, the El Farol Problem concerns the collective decision-making of a group of potential bar-goers (i.e. agents) who repeatedly try to predict whether they should attend a potentially overcrowded bar on a given night each week. They have no information about the others' predictions. Indeed the only information available to each agent is global, comprising a string of outcomes ('overcrowded' or 'undercrowded') for a limited number of previous occasions.



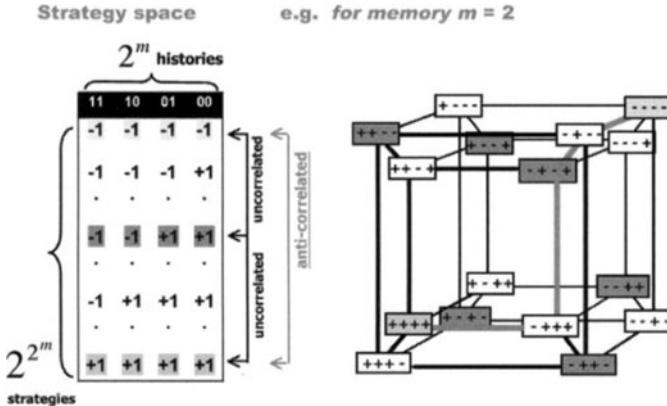
**Fig. 1.** Schematic representation of B-A-R (Binary Agent Resource) system. At timestep  $t$ , each agent decides between action  $-1$  and action  $+1$  based on the predictions of the  $S$  strategies in his possession. A total of  $n_{-1}[t]$  agents choose  $-1$ , and  $n_{+1}[t]$  choose  $+1$ . Agents may be subject to some underlying network structure which may be static or evolving, and ordered or disordered (see Refs. [7,8]). The agents' actions are aggregated, and a global outcome 0 or 1 is assigned. Strategies are rewarded/penalized one virtual point according to whether their predicted action would have been a winning/losing action

Hence they end up having to predict the predictions of others. No ‘typical’ agent exists, since all such typical agents would make the same decision, hence rendering their common prediction scheme useless. With the exception of Ref. [4], the physics literature has focused on a simplified binary form of the El Farol Problem called the Minority Game (MG) as introduced by Challet and Zhang [5,6].

In this paper, we present a theoretical framework for describing a class of Complex Adaptive Systems comprising competitive multi-agent populations, which we refer to as B-A-R (Binary Agent Resource) systems. The resulting Crowd-Anticrowd theory is not limited to MG-like games, even though we focus on MG-like games in order to demonstrate the accuracy of the approach. The theory is built around the correlations or ‘crowding’ in strategy-space. Since the theory only makes fairly modest assumptions about a specific game’s dynamical behavior, it can describe the dynamics in a wide variety of systems comprising competitive populations [7].

## 2 B-A-R (Binary Agent Resource) Systems

Figure 1 summarizes the generic form of the B-A-R (Binary Agent Resource) system under consideration. At timestep  $t$ , each agent (e.g. a bar customer,



**Fig. 2.** Strategy space for  $m = 2$ , together with some example strategies (left). The strategy space shown is known as the Full Strategy Space (FSS) and contains all possible permutations of the actions  $-1$  and  $+1$  for each history. There are  $2^{2^m}$  strategies in the FSS. The  $2^m$  dimensional hypercube (right) shows all  $2^{2^m}$  strategies in the FSS at its vertices. The shaded strategies form a Reduced Strategy Space RSS. There are  $2 \cdot 2^m = 2P$  strategies in the RSS. The grey shaded line connects two strategies with a Hamming distance separation of 4

a commuter, or a market agent) decides whether to enter a game where the choices are action  $+1$  (e.g. attend the bar, take route A, or buy) and action  $-1$  (e.g. go home, take route B, or sell). The global information available to the agents is a common memory of the most recent  $m$  outcomes, which are represented as either 0 (e.g. bar attendance below seating capacity  $L$ ) or 1 (e.g. bar attendance above seating capacity  $L$ ). Hence this outcome history is represented by a binary bit-string of length  $m$ . For general  $m$ , there will be  $P = 2^m$  possible history bit-strings. These history bit-strings can alternatively be represented in decimal form:  $\mu = \{0, 1, \dots, P - 1\}$ . For  $m = 2$ , for example,  $\mu = 0$  corresponds to 00,  $\mu = 1$  corresponds to 01 etc. A strategy consists of a predicted action,  $-1$  or  $+1$ , for each possible history bit-string. Hence there are  $2^{P=2^m}$  possible strategies.

Figure 2 shows the  $m = 2$  strategy space from Fig. 1. A strategy is a set of instructions to describe what action an agent should take, given any particular history  $\mu$ . The strategy space is the set of strategies from which agents are randomly allocated their strategies. The strategy space shown is known as the Full Strategy Space (FSS) and contains all possible permutations of the actions  $-1$  and  $+1$  for each history. As such there are  $2^{2^m}$  strategies in this space. One can choose a subset of  $2 \cdot 2^m$  strategies, called a Reduced Strategy Space (RSS), such that any pair within this subset has one of the following two characteristics: (i) *Anti-correlated*. For example, any two agents using the ( $m = 2$ ) strategies  $(-1, -1, +1, +1)$  and  $(+1, +1, -1, -1)$  respectively, would take the opposite action irrespective of the sequence of

previous outcomes and hence the history. Their net effect on the excess demand  $D[t] = n_{+1}[t] - n_{-1}[t]$  (which is an important quantity in a socio-economic setting such as a financial market) therefore cancels out at each timestep, regardless of the history. Hence they will not contribute to fluctuations in  $D[t]$ . (ii) *Uncorrelated*. For example, any two agents using the strategies  $(-1, -1, -1, -1)$  and  $(-1, -1, +1, +1)$  respectively, would take the opposite action for two of the four histories, and the same action for the remaining two histories. If the histories occur equally often, the actions of the two agents will be uncorrelated on average. Note that the strategies in the RSS can be labeled by  $R = \{1, 2, \dots, 2P = 2.2^m\}$ .

The strategy allocation among agents can be described in terms of a tensor  $\Omega$  describing the distribution of strategies among the  $N$  individual agents. This strategy allocation is typically fixed from the beginning of the game, hence acting as a quenched disorder in the system. The rank of the tensor  $\Omega$  is given by the number of strategies  $S$  that each agent holds. We note that a single  $\Omega$  ‘macrostate’ corresponds to many possible ‘microstates’ describing the specific partitions of strategies among the agents. Hence the present Crowd-Anticrowd theory retained at the level of a given  $\Omega$ , describes the set of all games which belong to that same  $\Omega$  macrostate. We also note that although  $\Omega$  is not symmetric, it can be made so since the agents do not distinguish between the order in which the two strategies are picked. Given this, we will henceforth focus on  $S = 2$  and consider the symmetrized version of the strategy allocation matrix given by  $\Psi = \frac{1}{2}(\Omega + \Omega^T)$ .

### 3 Crowd-Anticrowd Formalism

Consider an arbitrary timestep  $t$  during a run of the game. We will focus on evaluating the ‘excess demand’  $D[t] \equiv D[\underline{S}[t], \mu[t]] = n_{+1}[t] - n_{-1}[t]$ , although any other function of  $n_{+1}[t]$  and  $n_{-1}[t]$  can be evaluated in a similar way. Here  $\underline{S}[t]$  is the  $2P$ -dimensional score-vector whose  $R$ ’th component is the virtual point score for strategy  $R$ . [Strategies gain/lose one virtual point at each timestep, according to whether their predicted action would have been a winning/losing action]. The current history is  $\mu[t]$ . The standard deviation of  $D[t]$  for this given run, corresponds to a time-average for a given realization of  $\Psi$  and a given set of initial conditions. Summing over the RSS, we have:  $D[\underline{S}[t], \mu[t]] = \sum_{R=1}^{2P} a_R^{\mu[t]} n_R^{\underline{S}[t]}$ . The quantity  $a_R^{\mu[t]} = \pm 1$  is the action predicted by strategy  $R$  in response to the history bit-string  $\mu$  at time  $t$ . The quantity  $n_R^{\underline{S}[t]}$  is the number of agents using strategy  $R$  at time  $t$ . We use the notation  $\langle X[t] \rangle_t$  to denote a time-average over the variable  $X[t]$  for a given  $\Psi$ . Hence

$$\langle D[\underline{S}[t], \mu[t]] \rangle_t = \sum_{R=1}^{2P} \left\langle a_R^{\mu[t]} n_R^{\underline{S}[t]} \right\rangle_t = \sum_{R=1}^{2P} \left\langle a_R^{\mu[t]} \right\rangle_t \left\langle n_R^{\underline{S}[t]} \right\rangle_t \quad (1)$$

where we have used the property that  $a_R^{\mu[t]}$  and  $n_R^{S[t]}$  are uncorrelated. We now consider the special case in which all histories are visited equally on average: even if this situation does not hold for a specific  $\Psi$ , it may indeed hold once the averaging over  $\Psi$  has also been taken. For example, in the Minority Game all histories are visited equally at small  $m$  and a given  $\Psi$ . If we take the additional average over all  $\Psi$ , then the same is also true for large  $m$ . Under the property of equal histories:

$$\begin{aligned} \langle D[\underline{S}[t], \mu[t]] \rangle_t &= \sum_{R=1}^{2P} \left( \frac{1}{P} \sum_{\mu=0}^{P-1} a_R^{\mu[t]} \right) \langle n_R^{S[t]} \rangle_t \\ &= \sum_{R=1}^P \left( \frac{1}{P} \sum_{\mu=0}^{P-1} a_R^{\mu[t]} + a_{\bar{R}}^{\mu[t]} \right) \langle n_R^{S[t]} \rangle_t = \sum_{R=1}^P 0. \langle n_R^{S[t]} \rangle_t \\ &= 0 \end{aligned} \quad (2)$$

where we have used the exact result that  $a_R^{\mu[t]} = -a_{\bar{R}}^{\mu[t]}$  for all  $\mu[t]$ , and the approximation  $\langle n_R^{S[t]} \rangle_t = \langle n_{\bar{R}}^{S[t]} \rangle_t$ . This approximation is reasonable for a competitive game since there is typically no a priori best strategy: if the strategies are distributed fairly evenly among the agents, this then implies that the average number playing each strategy is approximately equal and hence  $\langle n_R^{S[t]} \rangle_t = \langle n_{\bar{R}}^{S[t]} \rangle_t$ . In the event that all histories are not equally visited over time, even after averaging over all  $\Psi$ , it may still happen that the system's dynamics is restricted to equal visits to some *subset* of histories. In this case one can then carry out the averaging in Eq. (2) over this subspace of histories. More generally, the averagings in this formalism can be carried out with appropriate frequency weightings for each history. In fact, any non-ergodic dynamics can be incorporated if one knows the appropriate history path [8].

The variance of the excess demand  $D[t]$  is given by

$$\sigma_\Psi^2 = \left\langle D[\underline{S}[t], \mu[t]]^2 \right\rangle_t - \langle D[\underline{S}[t], \mu[t]] \rangle_t^2. \quad (3)$$

For simplicity, we will here assume the game output is unbiased and hence  $\langle D[\underline{S}[t], \mu[t]] \rangle_t = 0$ . Hence

$$\sigma_\Psi^2 = \left\langle D[\underline{S}[t], \mu[t]]^2 \right\rangle_t = \sum_{R, R'=1}^{2P} \left\langle a_R^{\mu[t]} n_R^{S[t]} a_{R'}^{\mu[t]} n_{R'}^{S[t]} \right\rangle_t. \quad (4)$$

Using the identities  $\underline{a}_R \cdot \underline{a}_{R'} = P$  (fully correlated),  $\underline{a}_R \cdot \underline{a}_{R'} = -P$  (fully anti-correlated), and  $\underline{a}_R \cdot \underline{a}_{R'} = 0$  (fully uncorrelated), where  $\underline{a}_R$  is a vector of dimension  $P$  with components  $a_R^{\mu[t]}$  for  $\mu[t] = 1, 2, \dots, P$ , we obtain

$$\sigma_\Psi^2 = \sum_{R=1}^{2P} \left\langle \left( n_R^{S[t]} \right)^2 - n_R^{S[t]} n_{\bar{R}}^{S[t]} \right\rangle_t + \sum_{R \neq R' \neq \bar{R}}^{2P} \left\langle a_R^{\mu[t]} a_{R'}^{\mu[t]} \right\rangle_t \left\langle n_R^{S[t]} n_{R'}^{S[t]} \right\rangle_t$$

$$= \sum_{R=1}^{2P} \left\langle \left( n_R^{\underline{S}[t]} \right)^2 - n_R^{\underline{S}[t]} n_{\overline{R}}^{\underline{S}[t]} \right\rangle_t . \quad (5)$$

The sum over  $2P$  terms can be written equivalently as a sum over  $P$  terms,

$$\begin{aligned} \sigma_{\Psi}^2 &= \sum_{R=1}^P \left\langle \left( n_R^{\underline{S}[t]} \right)^2 - n_R^{\underline{S}[t]} n_{\overline{R}}^{\underline{S}[t]} + \left( n_{\overline{R}}^{\underline{S}[t]} \right)^2 - n_{\overline{R}}^{\underline{S}[t]} n_R^{\underline{S}[t]} \right\rangle_t \\ &= \sum_{R=1}^P \left\langle \left( n_R^{\underline{S}[t]} - n_{\overline{R}}^{\underline{S}[t]} \right)^2 \right\rangle_t \equiv \left\langle \sum_{R=1}^P \left( n_R^{\underline{S}[t]} - n_{\overline{R}}^{\underline{S}[t]} \right)^2 \right\rangle_t . \end{aligned}$$

The values of  $n_R^{\underline{S}[t]}$  and  $n_{\overline{R}}^{\underline{S}[t]}$  for each  $R$  will depend on the precise form of  $\Psi$ . The ensemble-average over all possible realizations of the strategy allocation matrix  $\Psi$  is denoted by  $\langle \dots \rangle_{\Psi}$ . Using the notation  $\langle \sigma_{\Psi}^2 \rangle_{\Psi} = \sigma^2$ , yields

$$\sigma^2 = \left\langle \left\langle \sum_{R=1}^P \left( n_R^{\underline{S}[t]} - n_{\overline{R}}^{\underline{S}[t]} \right)^2 \right\rangle_t \right\rangle_{\Psi} . \quad (6)$$

It is straightforward to obtain analogous expressions for the variances in  $n_{+1}[t]$  and  $n_{-1}[t]$ .

Equation (6) provides us with an expression for the time-averaged fluctuations. Some form of approximation must be introduced in order to reduce Eq. (6) to explicit analytic expressions. It turns out that Eq. (6) can be manipulated in a variety of ways, depending on the level of approximation that one is prepared to make. The precise form of any resulting analytic expression will depend on the details of the approximations made. Adopting one such approach which is well-suited to the low  $m$  regime, we start by relabelling the strategies. Specifically, the sum in Eq. (6) is re-written to be over a *virtual-point ranking*  $K$  as opposed to  $R$ . Consider the variation in points for a given strategy, as a function of time for a given realization of  $\Psi$ . The ranking (i.e. label) of a given strategy in terms of virtual-points score will typically change in time since the individual strategies have a variation in virtual-points which varies in time. For the Minority Game, this variation is quite rapid in the low  $m$  regime of interest, since there are many more agents than available strategies – hence any strategy emerging as the instantaneously highest-scoring, will immediately get played by many agents and therefore be likely to lose on the next time-step. More general games involving competition within a multi-agent population, will typically generate a similar ecology of strategy-scores with no all-time winner. This implies that the specific identity of the ' $K$ 'th highest-scoring strategy' changes frequently in time. It also implies that  $n_R^{\underline{S}[t]}$  varies considerably in time. Therefore in order to proceed, we shift the focus onto the time-evolution of the highest-scoring strategy, second highest-scoring strategy etc. This should have a much smoother time-evolution than the time-evolution for a given strategy. In the case that the strategies all

start off with zero points, the anticorrelated strategies appear as the mirror-image, i.e.  $S_K[t] = -S_{\bar{K}}[t]$ . The label  $K$  is used to denote the rank in terms of strategy score, i.e.  $\bar{K} = 1$  is the highest scoring strategy position,  $K = 2$  is the second highest-scoring strategy position etc. with

$$S_{K=1} > S_{K=2} > S_{K=3} > S_{K=4} > \dots \quad (7)$$

assuming no strategy-score ties. Given that  $S_R = -S_{\bar{R}}$  (i.e. all strategy scores start off at zero), then we know that  $S_K = -S_{\bar{K}}$ . Equation (6) can hence be rewritten exactly as

$$\sigma^2 = \left\langle \left\langle \sum_{K=1}^P \left( n_K^{\underline{S}[t]} - n_{\bar{K}}^{\underline{S}[t]} \right)^2 \right\rangle_t \right\rangle_{\Psi}. \quad (8)$$

Since in the systems of interest the agents are typically playing their highest-scoring strategies, then the relevant quantity in determining how many agents will instantaneously play a given strategy, is a knowledge of its relative ranking – not the actual value of its virtual points score. This suggests that the quantities  $n_K^{\underline{S}[t]}$  and  $n_{\bar{K}}^{\underline{S}[t]}$  will fluctuate relatively little in time, and that we should now develop the problem in terms of time-averaged values. We can rewrite the number of agents playing the strategy in position  $K$  at any timestep  $t$ , in terms of some constant value  $n_K$  plus a fluctuating term  $n_K^{\underline{S}[t]} = n_K + \varepsilon_K[t]$ . We assume that one can choose a suitable constant  $n_K$  such that the fluctuation  $\varepsilon_K[t]$  represents a small noise term. Hence,

$$\begin{aligned} \sigma^2 &= \left\langle \sum_{K=1}^P \left\langle [n_K + \varepsilon_K[t] - n_{\bar{K}} - \varepsilon_{\bar{K}}[t]]^2 \right\rangle_t \right\rangle_{\Psi} \\ &\approx \left\langle \sum_{K=1}^P \left\langle [n_K - n_{\bar{K}}]^2 \right\rangle_t \right\rangle_{\Psi} = \left\langle \sum_{K=1}^P [n_K - n_{\bar{K}}]^2 \right\rangle_{\Psi}, \end{aligned} \quad (9)$$

assuming the noise terms have averaged out to be small. The averaging over  $\Psi$  can now be taken inside the sum. Each term can then be rewritten exactly using the joint probability distribution for  $n_K$  and  $n_{\bar{K}}$ , which we shall call  $P(n_K, n_{\bar{K}})$ . Hence

$$\sigma^2 = \sum_{K=1}^P \left\langle [n_K - n_{\bar{K}}]^2 \right\rangle_{\Psi} = \sum_{K=1}^P \sum_{n_K=0}^N \sum_{n_{\bar{K}}=0}^N [n_K - n_{\bar{K}}]^2 P(n_K, n_{\bar{K}}). \quad (10)$$

We now look at Eq. (10) in the limiting case where the averaging over the quenched disorder matrix is dominated by matrices  $\Psi$  which are nearly flat. This will be a good approximation in the ‘crowded’ limit of small  $m$  in which there are many more agents than available strategies, since the standard deviation of an element in  $\Psi$  (i.e. the standard deviation in bin-size) is then much smaller than the mean bin-size. The probability distribution  $P(n_K, n_{\bar{K}})$

will then be sharply peaked around the  $n_K$  and  $n_{\bar{K}}$  values given by the mean values for a flat quenched-disorder matrix  $\Psi$ . We label these mean values as  $\bar{n}_K$  and  $\bar{n}_{\bar{K}}$ . Hence  $P(n_K, n_{\bar{K}}) = \delta_{n_K, \bar{n}_K} \delta_{n_{\bar{K}}, \bar{n}_{\bar{K}}}$  and so

$$\sigma^2 = \sum_{K=1}^P [\bar{n}_K - \bar{n}_{\bar{K}}]^2. \quad (11)$$

There is a very simple interpretation of Eq. (11). It represents the sum of the variances for each Crowd-Anticrowd pair. For a given strategy  $K$  there is an anticorrelated strategy  $\bar{K}$ . The  $\bar{n}_K$  agents using strategy  $K$  are doing the *opposite* of the  $\bar{n}_{\bar{K}}$  agents using strategy  $\bar{K}$  *irrespective* of the history bit-string. Hence the effective group-size for each Crowd-Anticrowd pair is  $n_K^{eff} = \bar{n}_K - \bar{n}_{\bar{K}}$  : this represents the net step-size  $d$  of the Crowd-Anticrowd pair in a random-walk contribution to the total variance. Hence, the net contribution by this Crowd-Anticrowd pair to the variance is given by

$$[\sigma^2]_{K\bar{K}} = 4pqd^2 = 4 \cdot \frac{1}{2} \cdot \frac{1}{2} [n_K^{eff}]^2 = [\bar{n}_K - \bar{n}_{\bar{K}}]^2 \quad (12)$$

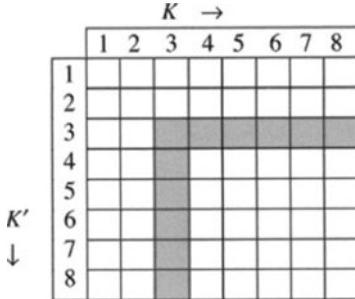
where  $p = q = 1/2$  for a random walk. Since all the strong correlations have been included (i.e. anti-correlations) it can therefore be assumed that the separate Crowd-Anticrowd pairs execute random walks which are *uncorrelated* with respect to each other. [Recall the properties of the RSS – all the remaining strategies are uncorrelated.] Hence the total variance is given by the sum of the individual variances,

$$\sigma^2 = \sum_{K=1}^P [\sigma^2]_{K\bar{K}} = \sum_{K=1}^P [\bar{n}_K - \bar{n}_{\bar{K}}]^2, \quad (13)$$

which corresponds exactly to Eq. (11). If strategy-ties occur frequently, then one has to be more careful about evaluating  $\bar{n}_K$  since its value may be affected by the tie-breaking rule. We will show elsewhere that this becomes quite important in the case of very small  $m$  in the presence of network connections [8] – this is because very small  $m$  naturally leads to crowding in strategy space and hence mean-reverting virtual scores for strategies. This mean-reversion is amplified further by the presence of network connections which increases the crowding, thereby increasing the chance of strategy ties.

## 4 Implementation of Crowd-Anticrowd Theory

Here we evaluate the Crowd-Anticrowd expressions, in the important limiting case of small  $m$ . Since there are many more agents than available strategies, crowding effects will be important. Each element of  $\Psi$  has a mean of  $N/(2P)^S$  agents per ‘bin’. In the case of small  $m$  and hence densely-filled  $\Psi$ , the fluctuations in the number of agents per bin will be small compared to this mean



**Fig. 3.** Schematic representation of the strategy allocation matrix  $\Psi$  with  $m = 2$  and  $S = 2$ , in the RSS. The strategies are ranked according to strategy score, and are labeled by the rank  $K$ . In the limit where  $\Psi$  is essentially flat, the number of agents playing the  $K$ 'th highest-scoring strategy, is just proportional to the number of shaded bins at that  $K$

value – hence the matrix  $\Psi$  looks uniform or ‘flat’ in terms of the occupation numbers in each bin. Figure 3 provides a schematic representation of  $\Psi$  with  $m = 2$ ,  $S = 2$ , in the RSS. If the matrix  $\Psi$  is indeed flat, then any re-ordering due to changes in the strategy ranking has no effect on the form of the matrix. Therefore the number of agents playing the  $K$ 'th highest-scoring strategy, will always be proportional to the number of shaded bins at that  $K$  (see Fig. 3 for  $K = 3$ ). For general  $m$  and  $S$ , one finds

$$\begin{aligned} \bar{n}_K &= \frac{N}{(2P)^S} [S(2P - K)^{S-1} + \frac{S(S-1)}{2}(2P - K)^{S-2} + \dots + 1] \quad (14) \\ &= \frac{N}{(2P)^S} \sum_{r=0}^{S-1} \frac{S!}{(S-r)!r!} [2P - K]^r \\ &= \frac{N}{(2P)^S} ([2P - K + 1]^S - [2P - K]^S) \\ &= N \cdot \left( \left[ 1 - \frac{(K-1)}{2P} \right]^S - \left[ 1 - \frac{K}{2P} \right]^S \right), \end{aligned}$$

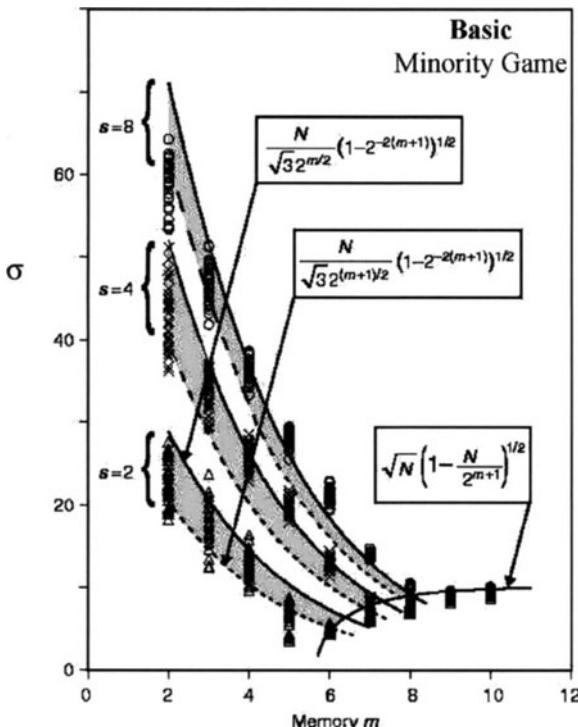
with  $P \equiv 2^m$ . In the case where each agent holds two strategies,  $S = 2$ ,  $\bar{n}_K$  can be simplified to

$$\bar{n}_K = N \cdot \left( \left[ 1 - \frac{(K-1)}{2P} \right]^2 - \left[ 1 - \frac{K}{2P} \right]^2 \right) = \frac{(2^{m+2} - 2K + 1)}{2^{2(m+1)}} N. \quad (15)$$

Hence

$$\begin{aligned} \sigma^2 &= \sum_{K=1}^P \left[ \frac{(2^{m+2} - 2K + 1)}{2^{2(m+1)}} N - \frac{(2K-1)}{2^{2(m+1)}} N \right]^2 \quad (16) \\ &= \frac{N^2}{2^{2(2m+1)}} \sum_{K=1}^P [2^{m+1} - 2K + 1]^2 = \frac{N^2}{3 \cdot 2^m} (1 - 2^{-2(m+1)}) . \end{aligned}$$

This derivation has assumed that there are no strategy ties – more precisely, we have assumed that the game rules governing strategy ties do not upset the



**Fig. 4.** Crowd-Anticrowd theory vs. numerical simulation results for  $\sigma$  in the Minority Game as a function of memory size  $m$ , for  $N = 101$  agents, at  $S = 2, 4$  and  $8$ . At each  $S$  value, analytic forms of  $\sigma$  (i.e. standard deviation in excess demand  $D[t]$ ) are shown. The numerical values were obtained from different simulation runs (triangles, crosses and circles). Figure adapted from Ref. [2]

identical forms of the rankings in terms of highest virtual points and popularity. Hence we have overestimated the size of the crowds using high-ranking strategies, and underestimated the size of the anticrowds using low-ranking strategies. Therefore the analytic form for  $\sigma$  will overestimate the numerical value, as is indeed seen in Fig. 4. Notwithstanding this overestimation, there is remarkably good agreement between the numerical results and our analytic theory. In a similar way to the above calculation, the Crowd-Anticrowd theory can be extended to deal with the important complementary regimes of (i) non-flat quenched disorder matrix  $\Psi$ , at small  $m$ , and (ii) non-flat quenched disorder matrix  $\Psi$ , at large  $m$ . As shown in Fig. 4, the agreement for these regimes is also excellent [2,7].

The Crowd-Anticrowd theory has also been applied successfully to various generalizations of the Minority Game. For example, excellent agreement between the resulting analytic expressions and numerical simulations has been demonstrated for (i) Alloy Minority Game [9], (ii) Thermal Minority Game

(TMG) [10,11], (iii) Thermal Alloy Minority Game [12], and (iv) B-A-R systems with an underlying network structure [7].

## 5 Conclusion and Discussions

We have given an overview of the Crowd-Anticrowd theory for competitive multi-agent systems, in particular those based on an underlying binary structure. Explicit analytic expressions can be evaluated at various levels of approximation, yielding very good agreement with numerical simulations. We note that the crucial element of this Crowd-Anticrowd theory – i.e. properly accounting for the dominant inter-agent correlations – is not limited to one specific game. Given its success in describing a number of generalized B-A-R systems, we believe that the Crowd-Anticrowd framework could provide a powerful approach to describing a wide class of Complex Adaptive Systems which mimic competitive multi-agent games. This would be a welcome development, given the lack of general theoretical concepts in the field of Complex Adaptive Systems as a whole. It is also pleasing from the point of view of physics methodology, since the basic underlying philosophy of accounting correctly for ‘inter-particle’ correlations is already known to be successful in more conventional areas of many-body physics. This success in turn raises the intriguing possibility that conventional many-body physics might be open to re-interpretation in terms of an appropriate multi-particle ‘game’: we leave this for future work.

Of course, some properties of Complex Adaptive Systems cannot be described using time- and configuration-averaged expressions as discussed here. In particular, an observation of a real-world Complex Adaptive System which is thought to resemble a multi-agent game, may correspond to a *single* run which evolves from a specific initial configuration of agents’ strategies. This implies a particular  $\Psi$ , and hence the time-averagings within the Crowd-Anticrowd theory must be carried out for that particular choice of  $\Psi$ . However this problem can still be cast in terms of the Crowd-Anticrowd approach, since the averagings are then just carried out over some sub-set of paths in history space, which is conditional on the path along which the Complex Adaptive System is already heading.

We have been discussing a Complex Adaptive System based on multi-agent dynamics, in which both deterministic and stochastic processes co-exist, and are indeed intertwined. Depending on the particular rules of the game, the stochastic element may be associated with any of five areas: (i) disorder associated with the strategy allocation and hence with the heterogeneity in the population, (ii) disorder in an underlying network. Both (i) and (ii) might typically be fixed from the outset (i.e., quenched disorder) hence it is interesting to see the interplay of (i) and (ii) in terms of the overall performance of the system [8]. The extent to which these two ‘hard-wired’ disorders might then compensate each other, as for example in the Parrondo effect or stochas-

tic resonance, is an interesting question. Such a compensation effect might be engineered, for example, by altering the rules-of-the-game concerning inter-agent communication on the existing network. Three further possible sources of stochasticity are (iii) tie-breaks in the scores of strategies, (iv) a stochastic rule in order for each agent to pick which strategy to use from the available  $S$  strategies, as in the Thermal Minority Game, (v) stochasticity in the global resource level  $L[t]$  (e.g. bar seating capacity) due to changing external conditions. To a greater or lesser extent, these five stochastic elements will tend to break up any deterministic cycles arising in the game. We refer to Ref. [13] for a discussion of the dynamics of the Minority Game viewed from the perspective of a stochastically-perturbed deterministic system.

## References

1. J.L. Casti, *Would-be Worlds* (Wiley, New York, 1997).
2. N.F. Johnson, P. Jefferies, P.M. Hui, *Financial Market Complexity* (Oxford University Press, 2003).
3. W.B. Arthur, Amer. Econ. Rev. **84**, 406 (1994); Science **284**, 107 (1999).
4. N.F. Johnson, S. Jarvis, R. Jonson, P. Cheung, Y. Kwong and P.M. Hui, Physica A **258**, 230 (1998).
5. D. Challet and Y.C. Zhang, Physica A **246**, 407 (1997).
6. See <http://www.unifr.ch/econophysics/minority> and E. Moro e-print cond-mat/0402651 at xxx.lanl.gov for the full MG-related literature.
7. See N.F. Johnson and P.M. Hui, e-print cond-mat/0306516 at xxx.lanl.gov, for more details.
8. S. Gourley, S.C. Choe, P.M. Hui and N.F. Johnson, e-print cond-mat/0401526 at xxx.lanl.gov.
9. N.F. Johnson, P.M. Hui, Dafang Zheng, and M. Hart, J. Phys. A: Math. Gen. **32**, L427 (1999).
10. A. Cavagna, J.P. Garrahan, I. Giardina and D. Sherrington, Phys. Rev. Lett. **83**, 4429 (1999).
11. M.L. Hart, P. Jefferies, N.F. Johnson and P.M. Hui, Phys. Rev. E **63**, 017102 (2001).
12. P. Jefferies, M. Hart, N.F. Johnson, and P.M. Hui, J. Phys. A: Math. Gen. **33**, L409 (2000).
13. P. Jefferies, M.L. Hart and N.F. Johnson, Phys. Rev. E **65**, 016105 (2002).

# Modified Mode Coupling Theory of Glassy Dynamics Generated by Entanglement

Rolf Schilling

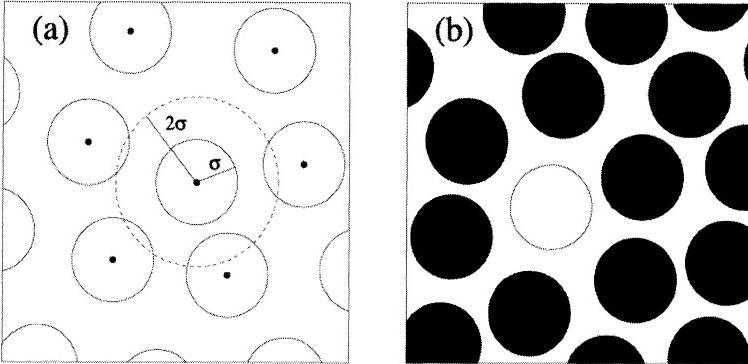
Institut für Physik, Johannes Gutenberg-Universität Mainz  
Staudinger Weg 7, 55099 Mainz, Germany

**Abstract.** Increasing the density in systems with strong excluded volume interactions leads simultaneously to an increase of static correlations and a slowing down of the relaxational dynamics. Mode coupling theory in its present form describes this mechanism, satisfactorily. In contrast, for systems where entanglement is dominant, e.g. infinitely thin hard rods on a lattice, glassy dynamics is not driven by increasing static correlations but by entanglement. We show how mode coupling approximation can be modified such that non-vanishing vertices occur which might account for such pure entanglement effects.

## 1 Introduction

The static and dynamical properties of all classical liquids are governed by *excluded volume* effects. The most prominent case is a system of hard spheres with radius  $\sigma$  and number density  $n$ . With each sphere one can associate a fictitious sphere of radius  $2\sigma$  (Fig. 1a) which can not be penetrated by the center of any other sphere. This is the excluded volume effect. If  $n$  is rather small the static correlations are weak and the structural rearrangement takes place on a short time scale. If transitions to ordered phases do not occur, increasing  $n$  results in a strong increase of the static correlations which is accompanied by a slowing down of the structural relaxation. This behavior can be quantified by use of the intermediate scattering function  $S(q, t)$ , for e.g. a simple liquid.  $S(q, t)$  is the time dependent correlation function of the density fluctuations with wave number  $q = |\mathbf{q}|$ . The *static* correlator is given by  $S(q) \equiv S(q, 0)$ . With increasing  $n$  the peaks of  $S(q)$  and particularly its main peak, which is related to nearest neighbor correlations, increase. Simultaneously the decay of  $S(q, t)$  with time slows down. Both phenomena are strongly interrelated and their origin is the so-called *cage effect*, occurring due to the excluded volume interactions (Fig. 1b). Increasing  $n$  even more may lead to a structural arrest at a critical density  $n_c$ . In that case the low temperature phase is a glass and the transition is the structural glass transition.

There exist a couple of phenomenological theories of glassy dynamics and the glass transition (see e.g. the reviews [1,2]), but they do not have much predictive power. It has taken many decades till a theory, the *mode coupling*



**Fig. 1.** Schematic illustration of hard spheres (disks) in two dimensions. (a) low density  $n$ : the dashed line indicates the fictitious sphere which can not be penetrated by the center of any other sphere. (b) high density  $n$ : a marked sphere (open circle) in a cage of other spheres (full circle)

*theory* (MCT), was derived by Götze and his coworkers [3] being capable to describe the interplay between the growth of static correlations and the slowing down of the structural relaxation on a *microscopic level*. For more details see Refs. [4,5,6,7,8,9].

One may ask whether excluded volume interactions are necessary for glassy dynamics and the glass transition. The answer is no! Many models were suggested and investigated which exhibit slowing down of the dynamics without excluding volume effects (see the recent review [10]). One of these models are infinitely thin hard rods of length  $L = l \cdot a$  on a periodic lattice with lattice constant  $a$ . Since the rod's diameter  $d$  is zero there are no excluded volume interactions and the static correlations vanish. Nevertheless, the rods start to *entangle* for  $l > 1$ . The entanglement increases with increasing  $l$ . This leads to a dynamical cage effect which affects the orientational relaxation. That this really happens has been demonstrated by a MD-[11] and MC-simulation [12]. In particular it has been shown [11] that the rotational diffusion constant  $D(l)$  for rods fixed with their center of mass on a fcc-lattice can be fitted by a power law  $(l_c - l)^\gamma$  with  $l_c \approx 2.7$  and  $\gamma \approx 4.2$ . In the idealized situation an orientational glass transition would take place at the critical length  $l_c$ . Furthermore, this transition is discontinuous [11]. This means that the long time limit of an orientational correlator, the so-called non-ergodicity parameter, changes discontinuously from zero in the ergodic phase to a nonzero value in the glass phase. These findings are similar to the universal features of glassy dynamics as predicted by conventional MCT. However, they can not be treated in the framework of the microscopic version of MCT, because the vertices vanish.

Another example are polymeric systems. There already exist MCT-approaches for dense polymer melts accounting for entanglement (see e.g. [13,14,15,16,17] and references therein). But *all* these theoretical at-

tempts assume the equilibrium structure as the origin of entanglement formation. Now, idealizing a polymer chain as an infinitely thin thread with finite stiffness and assuming hard body interactions between different chains, the inter-chain dynamics will be still governed by entanglement, although the equilibrium structure becomes trivial, i.e. the direct correlation function vanishes. It is such an idealized situation like the hard rods on a lattice or infinitely thin treads in a liquid which we have in mind. It is of fundamental and practical interest to understand the effect of entanglement on glassy dynamics.

In a first attempt we have worked out a microscopic theory [18,19] for the system of rods. The starting point is the Smoluchowski equation for the time-dependent  $N$ -rod probability density  $P_N(\Omega_1, \dots, \Omega_N; t)$ .  $\Omega_i = (\Theta_i, \phi_i)$  describes the orientation of the rod at lattice site  $i$ . Then it is straightforward to derive a hierarchy of equations for the reduced  $k$ -rod density

$$\rho_{n_1 \dots n_k}^{(k)}(\Omega_1, \dots, \Omega_k; t) = \int \prod_{n \neq n_1, \dots, n_k} d\Omega_n P_N(\Omega_1, \dots, \Omega_N; t). \quad (1)$$

where we use  $\Omega_i \equiv \Omega_{n_i}$  on the l.h.s. of Eq. (1). This hierarchy relates  $\partial \rho_{n_1}^{(1)} / \partial t$  with  $\rho_{n_1}^{(1)}$  and  $\rho_{n_1 n_2}^{(1)}$  and  $\partial \rho_{n_1 n_2}^{(2)} / \partial t$  with  $\rho_{n_1}^{(1)}$ ,  $\rho_{n_1 n_2}^{(2)}$  and  $\rho_{n_1 n_2 n_3}^{(3)}$ , etc. In order to close these equations we neglect  $\rho_{n_1 n_2 n_3}^{(3)}$  and describe the influence of a third rod at  $n_3$  by replacing the bare diffusivity tensor  $\mathbf{D}_0$  by an effective one,  $\widehat{\mathbf{D}}(\Omega_1, \Omega_2; t)$ .

Finally one arrives at a self-consistency equation for the Laplace transform  $\widehat{\mathbf{D}}(\Omega_1, \Omega_2; z)$  [19]

$$\begin{aligned} \widehat{\mathbf{D}}^{-1}(\Omega_1, \Omega_2; z) &= \mathbf{D}_0^{-1} \delta(\Omega_1 | \Omega_2) + \\ &+ \frac{1}{4\pi} \sum_{n_3} \langle \mathbf{T}_{n_1 n_3}(\Omega_1, \Omega_2) [(z - \widehat{\Lambda}^{(2)})^{-1} \\ &\quad * (\mathbf{T}_{n_1 n_3}^t g_{n_1 n_3}^{(2)})]_{n_1 n_3}(\Omega_1, \Omega_2; z) \rangle \end{aligned} \quad (2)$$

where  $(\mathbf{D}_0)_{\alpha\beta} = D_0 \delta_{\alpha\beta}$ ,  $g_{n_1 n_2}^{(2)}(\Omega_1, \Omega_2)$  the equilibrium pair distribution function,  $\widehat{\Lambda}^{(2)}$  a linear operator [19] and  $\mathbf{T}_{n_1 n_2}(\Omega_1, \Omega_2)$  is the torque between rods at site  $n_1$  and  $n_2$ . For technical convenience we perform a further approximation

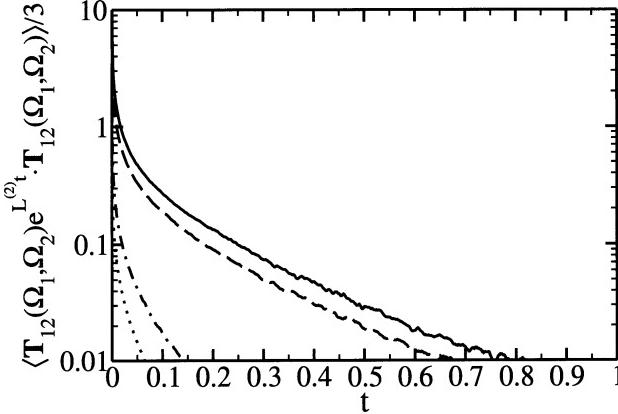
$$\widehat{\mathbf{D}}(\Omega_1, \Omega_2; z))_{\alpha\beta} \approx \widehat{D}(z) \delta_{\alpha\beta} \delta(\Omega_1 | \Omega_2). \quad (3)$$

Using Eq. (3) one obtains from Eq. (2) [19],

$$\frac{1}{\widehat{D}(z)} = \frac{1}{D_0} + \frac{1}{\widehat{D}(z)} \widehat{v}(\zeta; l), \quad \zeta = z/\widehat{D}(z), \quad (4)$$

where  $\widehat{v}(\zeta; l)$  is related to the Laplace transform of the torque-torque correlator

$$v(t; l) = \frac{1}{3} \sum_{n_2} \langle \mathbf{T}_{n_1 n_2} \cdot e^{\mathcal{L}^{(2)\dagger} t} \mathbf{T}_{n_1 n_2} \rangle. \quad (5)$$



**Fig. 2.** Time-dependence of the torque-torque correlator for  $L/|\mathbf{r}_{n_1} - \mathbf{r}_{n_2}| = 1.8$  (dotted), 2.0 (dash-dotted), 4 (dashed) and 6 (solid line).  $\mathbf{r}_n$  is the lattice vector of site  $n$

$\langle \quad \rangle$  denotes averaging over  $\Omega_1$  and  $\Omega_2$  with  $(1/4\pi)^2 g_{n_1 n_2}^{(2)}(\Omega_1, \Omega_2)$  as measure and  $\mathcal{L}^{(2)\dagger}$  is the adjoint 2-rod Smoluchowski operator with  $D_0$  replaced by one. From Eq. (4) one gets immediately the corresponding equation for the diffusion constant  $D(l)$  at  $z = 0$

$$D(l) = D_0[1 - v(l)] \quad (6)$$

with

$$v(l) \equiv \hat{v}(0; l) = \int_0^\infty dt v(t; l). \quad (7)$$

This result shows that the bare diffusion constant is reduced by a factor  $(1 - v(l))$ . The crucial quantity  $v(l)$  is determined by the Laplace transform at frequency  $z = 0$  of the torque-torque correlator  $\langle \mathbf{T}_{n_1 n_2} \cdot \exp \mathcal{L}^{(2)\dagger} t \mathbf{T}_{n_1 n_2} \rangle$  of an *isolated* pair of rods at sites  $n_1$  and  $n_2$ . A numerical result for that correlator is presented in Fig. 2. Figure 2 demonstrates that the decay of the torque-torque correlator becomes slower with increasing  $l$ . Therefore the integral in Eq. (7) and accordingly  $v(l)$  increases. An orientational arrest occurs if  $D(l) = 0$  which happens at a critical length  $l_c$  for which  $v(l_c) = 1$ . Making use of the numerical results for  $v(l)$  we have got  $l_c \cong 3.45$ , 2.78 and 2.20 for a sc, bcc and fcc lattice, respectively. The value  $l_c \cong 2.20$  is in a reasonable range of  $l_c^{\text{MD}} \approx 2.7$  [11]. A variational approach for the calculation of  $v(l)$  yields values for  $l_c$  which exceed the numerical ones by roughly 50% [19]. It is easy to show that  $D(l) \sim (l_c - l)^\gamma$  with  $\gamma = 1$ .

Although we have found a self-consistency equation which indeed yields a glass transition with reasonable values for  $l_c$  it can not be the correct one. A discussion of the  $z$ -dependent self-consistency equation Eq. (4) allows to

prove that the non-ergodicity parameters are continuous at  $l_c$  [19], in variance with the numerical finding. Consequently we have to improve our approach, qualitatively, in order to remove that drawback. How this can be achieved is the subject of the next section.

## 2 Microscopic Theory of a Discontinuous Glass Transition

Since we have not found a way to modify the effective medium-like theory presented in Sect. 1 such that a discontinuous glass transition occurs, we have returned to the Zwanzig-Mori formalism. We restrict ourselves to the system of  $N$  hard rods. Particles in a liquid can be treated similarly. We will consider  $N$  infinitely thin hard rods of length  $L$  on a periodic lattice with lattice constant  $a$ . The dynamics is Newtonian. The rods collide elastically, provided  $l = L/a > 1$  and move freely between subsequent collisions. For simplicity we assume the rods to be fixed with their center of mass. The corresponding moment of inertia is denoted by  $I$ . The coordinates in phase space are the angular coordinates  $\Omega_n = (\Theta_n, \phi_n)$  and their conjugate momenta  $p_{\Theta_n} = I\dot{\Theta}_n, p_{\phi_n} = I \sin^2 \Theta_n \cdot \dot{\phi}_n, n = 1, \dots, N$ . The kinetic energy is given by

$$\begin{aligned} \mathcal{H}_{\text{kin}}(\{\Theta_n, \phi_n\}, \{p_{\Theta_n}, p_{\phi_n}\}) &= \frac{1}{2I} \sum_{n=1}^N \left[ p_{\Theta_n}^2 + \sin^{-2} \Theta_n p_{\phi_n}^2 \right] \\ &= \frac{1}{2I} \sum_{n=1}^N \mathbf{l}_n^2 \end{aligned} \quad (8)$$

where  $\mathbf{l}_n$  is the classical angular momentum of rod  $n$ . The time evolution is determined by the Pseudo-Liouvillean

$$\mathcal{L}_{\pm} = \mathcal{L}_0 + \mathcal{L}'_{\pm} \quad (9)$$

$$\mathcal{L}_0 = -i\{\mathcal{H}_{\text{kin}}, \quad \} \quad (10)$$

$$\mathcal{L}'_{\pm} = -\frac{i}{2} \sum_{n \neq m} T_{nm}^{\pm} \quad (11)$$

with

$$\begin{aligned} T_{nm}^{\pm} &= \left| \frac{d}{dt} \mathbf{r}_{nm}^{\perp} \right| \Theta \left( \pm \frac{d}{dt} |\mathbf{r}_{nm}^{\perp}| \right) \Theta \left( \frac{L}{2} - |s_{nm}| \right) \\ &\times \Theta \left( \frac{L}{2} - |s_{mn}| \right) \delta \left( |\mathbf{r}_{nm}^{\perp}| - 0^+ \right) \left( b_{nm}^{\pm} - 1 \right). \end{aligned} \quad (12)$$

and  $b_{nm}^+(b_{nm}^-)$  replaces the angular momenta before (after) the collision by the corresponding momenta after (before) the collision. More details about the derivation of Eqs. (12) can be found in Refs. [11,20]. The torque  $T_{nm}^\pm$  (Eq. (12)) for Newtonian dynamics differs from that for Brownian dynamics although the product of theta- and delta functions describing the condition for a contact are identical.

Due to the rod's zero thickness the partition function is given by

$$Z(T, N) = (8\pi^2 k_B T I)^N. \quad (13)$$

Accordingly the free energy is not singular and therefore no equilibrium phase transition exists. Next we introduce the *microscopic* orientational density

$$\rho_n(\Omega, t) = \delta(\Omega | \Omega_n(t)) \quad (14)$$

for the rod at site  $n$  at time  $t$ . It obeys the continuity equation

$$\dot{\rho}_n(\Omega, t) + i\widehat{\mathbf{L}} \cdot \mathbf{j}_n(\Omega, t) = 0 \quad (15)$$

with  $\widehat{\mathbf{L}}$  the operator of angular momentum acting on  $\Omega$  and

$$\mathbf{j}_n(\Omega, t) = \boldsymbol{\omega}_n(t) \delta(\Omega | \Omega_n(t)) \equiv \boldsymbol{\omega}_n(t) \rho_n(\Omega, t) \quad (16)$$

the orientational current density. The angular velocity follows from

$$\boldsymbol{\omega}_n(t) = \mathbf{l}_n(t)/I. \quad (17)$$

It is useful to introduce tensorial densities by expanding a local function  $F_n(\Omega)$  with respect to spherical harmonics,

$$F_{n\lambda} = \sqrt{4\pi} i^j \int d\Omega F_n(\Omega) Y_\lambda(\Omega) \quad , \quad \lambda = (j, m). \quad (18)$$

The inverse is given by

$$F_n(\Omega) = \frac{1}{\sqrt{4\pi}} \sum_{\lambda} (-i)^j F_{n\lambda} Y_{\lambda}^*(\Omega). \quad (19)$$

This yields

$$\rho_{n\lambda}(t) = \sqrt{4\pi} i^j Y_{\lambda}(\Omega_n(t)) \quad (20)$$

$$\mathbf{j}_{n\lambda}(t) = \sqrt{4\pi} i^j \boldsymbol{\omega}_n(t) Y_{\lambda}(\Omega_n(t)) \equiv I^{-1} \mathbf{l}_n(t) \rho_{n\lambda}(t) \quad (21)$$

where we used Eq. (17). Note, that (i)  $\rho_{n00}(t) \equiv 1$ , due to the normalization of  $\rho_n(\Omega, t)$  and (ii)  $\rho_{n\lambda}(t) \equiv 0$  for  $j$  odd, because of the rod's head-tail symmetry.

Now we can introduce the time dependent orientational correlators

$$S_{nn',\lambda\lambda'}(t) = \langle \rho_{n\lambda}(t)^* \rho_{n'\lambda'}(0) \rangle \equiv \langle \rho_{n\lambda}^* e^{-i\mathcal{L}_- t} \rho_{n'\lambda'} \rangle \quad (22)$$

$$J_{nn',\lambda\lambda'}^{\nu\nu'}(t) = \langle j_{n\lambda}^\nu(t)^* j_{n'\lambda'}^{\nu'}(0) \rangle \equiv \langle j_{n\lambda}^{\nu*} e^{-i\mathcal{L}_- t} j_{n'\lambda'}^{\nu'} \rangle. \quad (23)$$

Since  $j = 0$  does not describe orientational degrees of freedom it will be  $j \geq 1$  in the following. It is easy to show that the *static* correlators, i.e. the correlators at  $t = 0$ , are given by

$$S_{nn',\lambda\lambda'} = \delta_{nn'} \delta_{\lambda\lambda'}, \quad J_{nn',\lambda\lambda'}^{\nu\nu'} = (k_B T / I) j(j+1) \delta_{nn'} \delta_{\lambda\lambda'} \delta_{\nu\nu'}. \quad (24)$$

The time dependent and static correlators also depend on the physical parameters  $l, I$  and temperature  $T$  (cf. Eq. (24), for  $t = 0$ ). Although  $I$  depends on the rod's mass and  $l$  we will take it as a parameter. Then all the quantities can be scaled such that the  $I$ - and  $T$ -dependence drops out. Let us introduce the microscopic time

$$\tau = [I/(k_B T)]^{1/2}. \quad (25)$$

Then we will use the scaled variables

$$t \rightarrow t/\tau, \quad \mathcal{L}_\pm \rightarrow \mathcal{L}_\pm \tau, \quad \ell_n^\alpha \rightarrow (\tau/I) \ell_n^\alpha \quad (26)$$

from now on.

To apply the Zwanzig-Mori projection formalism we choose the projectors on  $\rho_{n\lambda}$  and  $\dot{\rho}_{n\lambda}$ . The resulting Mori-Zwanzig equation reads

$$\ddot{\mathbf{S}}(t) + \nu \dot{\mathbf{S}}(t) + \Omega^2 \mathbf{S}(t) + \Omega^2 \int_0^t dt' \mathbf{m}(t-t') \dot{\mathbf{S}}(t') = 0 \quad (27)$$

where we used the matrix notation  $\mathbf{S}(t) = (S_{nn',\lambda\lambda'}(t))$ . Furthermore it is

$$\nu_{nn',\lambda\lambda'} = \frac{i}{j'(j'+1)} \langle \dot{\rho}_{n\lambda}^* | \mathcal{L}_- | \dot{\rho}_{n'\lambda'} \rangle \quad (28)$$

the scaled *Enskog term*

$$\Omega_{nn',\lambda\lambda'} = \Omega_j \delta_{nn'} \delta_{\lambda\lambda'}, \quad \Omega_j = [j(j+1)]^{1/2} \quad (29)$$

the scaled *microscopic frequencies* and

$$m_{nn',\lambda\lambda'}(t) = [j(j+1)j'(j'+1)]^{-1} \langle \dot{\rho}_{n\lambda}^* | \mathcal{L}_- Q e^{-iQ\mathcal{L}_- Qt} Q \mathcal{L}_- | \dot{\rho}_{n'\lambda'} \rangle \quad (30)$$

the corresponding *memory kernel*.  $Q = 1 - P_\rho - P_j$  projects perpendicular to  $\rho_{n\lambda}$  and  $\dot{\rho}_{n\lambda}$ . These results are in close relationship to those for simple

liquids [4,5,6]. Note that Eqs. (27) and (30) do not contain any  $I$  and  $T$  dependence, due to the use of scaling Eq. (26).

As well known, it is the main problem to find an approximation for the memory kernel. The strategy of MCT in its original form is to project the fluctuating force  $Q\mathcal{L}_-\dot{\rho}_{n\lambda}\rangle$  onto a pair  $\rho_{n'\lambda'}\rho_{n''\lambda''}$  of density modes. It is easy to prove that this projection is zero because of the trivial statics. This result also holds by projecting on any product  $\rho_{n_1\lambda_1}\rho_{n_2\lambda_2}\cdots\rho_{n_\nu\lambda_\nu}$ . Consequently, we have to follow a different path.  $\mathcal{L}_-\dot{\rho}_{n\lambda} = \mathcal{L}_0\dot{\rho}_{n\lambda} + \mathcal{L}'_-\dot{\rho}_{n\lambda}$  consists of a kinetic part (first term) and a force term (second term). Since the slowing down of the rod's dynamics is assumed to be governed by the forces we perform the *first approximation*

$$\mathcal{L}_-\dot{\rho}_{n\lambda} \approx \mathcal{L}'_-\dot{\rho}_{n\lambda} = - \sum_{\lambda',\alpha} L_{\lambda\lambda'}^\alpha \sum_{n' \neq n} T_{n'n}^- l_n^\alpha \rho_{n\lambda'} \quad (31)$$

where we used Eqs. (11) and (15)-(17).

The matrices  $(L_{\lambda\lambda'}^\alpha)$  are given by  $L_{\lambda\lambda'}^\alpha = \int d\Omega Y_\lambda^*(\Omega) \hat{L}^\alpha Y_{\lambda'}(\Omega)$ . Making use of Eq. (31) we get for the Laplace transform of the kernel

$$\begin{aligned} \hat{m}_{nn',\lambda\lambda'}(z) &\approx [j(j+1)j'(j'+1)]^{-1} \sum_{\alpha,\alpha'} \sum_{\lambda'',\lambda'''} L_{\lambda\lambda''}^{\alpha*} L_{\lambda'\lambda''''}^{\alpha'} \\ &\cdot \sum_{n'',n'''} \langle \rho_{n\lambda''}^* l_n^\alpha T_{nn''}^- | Q[Q\mathcal{L}_-Q - z]^{-1} Q | T_{n'''n'}^- l_{n'}^{\alpha'} \rho_{n'\lambda''''} \rangle. \end{aligned} \quad (32)$$

Note that  $T_{nn'}^-$  acts on  $l_n^\alpha$ , only. The *second approximation* is crucial

$$\begin{aligned} \langle \rho_{n\lambda''}^* l_n^\alpha T_{nn''}^- | Q[Q\mathcal{L}_-Q - z]^{-1} Q | T_{n'''n'}^- l_{n'}^{\alpha'} \rho_{n'\lambda''''} \rangle &\approx \\ \left\langle \rho_{n\lambda''}^* \left\{ \sum_{\lambda_1} \rho_{n\lambda_1} \rho_{n''\lambda_1}^* \right\} \left| Q[Q\mathcal{L}_-Q - z]^{-1} Q \right| \left\{ \sum_{\lambda'_1} \rho_{n'''n'}^* \rho_{n'\lambda'_1} \right\} \rho_{n'\lambda''''} \right\rangle \\ \times (-i) \left\langle l_n^\alpha \left| T_{nn''}^- \left[ \mathcal{L}^{(nn''n''n')} - z \right]^{-1} T_{n'''n'}^- \right| l_{n'}^{\alpha'} \right\rangle, (z=0). \end{aligned} \quad (33)$$

Here a few comments are in order. First, we have factorized the l.h.s. of Eq. (33) into a correlation function involving products of densities, only, and another one depending on the “forces”, e.g.  $T_{n'''n'}^- l_{n'}^{\alpha'}$ , only. Second, in order to account for the correct statistical weight of the force mediated by  $T_{nn''}^-$  and  $T_{n'''n'}^-$  we have introduced the terms in curly brackets. The sum  $\sum_{\lambda'_1}$  has the correct symmetry properties with respect to rotations and it is restricted to  $j'_1$  even (denoted by  $\sum'$ ), due to the head tail symmetry and  $n \neq n''$ ,  $n' \neq n'''$ . Third, in analogy to conventional MCT we assume the second factor not to be singular. This is achieved by restricting the time development to that of the rods at  $n, n'', n'''$  and  $n'$ , given by the corresponding Liouvillean  $\mathcal{L}^{(nn''n''n')}$ . Since this “force-force”-correlator (or more precisely torque-torque correlator) does not slow down, we are allowed to take it at frequency  $z = 0$ , provided the first factor exhibits slowing down.

Application of the product rule for the spherical harmonics on  $\rho_{n' \lambda'_1}^* \rho_{n' \lambda'''} \rho_{n' \lambda'_2}$  gives

$$\rho_{n' \lambda'_1}^* \rho_{n' \lambda'''} = \sum_{\lambda'_2} b_{\lambda'_1 \lambda'_2 \lambda''' \lambda'_2} \rho_{n' \lambda'_2} \quad (34)$$

with

$$\begin{aligned} b_{\lambda_1 \lambda_2 \lambda_3} = & \\ & i^{j_1 + j_2 - j_3} \left[ \frac{(2j_1 + 1)(2j_2 + 1)}{2j_3 + 1} \right]^{1/2} \\ & \times C(j_1 j_2 j_3 | (m_1 m_2 m_3) C(j_1 j_2 j_3 | 000) \equiv b_{\lambda_2 \lambda_1 \lambda_3}. \end{aligned} \quad (35)$$

The first factor can be expressed as linear combination of

$$\left\langle \rho_{n \lambda_2}^* \rho_{n'' \lambda_1}^* \left| Q [Q \mathcal{L}_- Q - z]^{-1} Q \right| \rho_{n''' \lambda'_1} \rho_{n' \lambda'_2} \right\rangle. \quad (36)$$

This correlation function of *density pairs* also appears in conventional MCT. Following MCT, we use as *third approximation*

$$\begin{aligned} & \left\langle \rho_{n \lambda_2}^* \rho_{n'' \lambda_1}^* \left| Q e^{-iQ\mathcal{L}_- Qt} Q \right| \rho_{n''' \lambda'_1} \rho_{n' \lambda'_2} \right\rangle \\ & \approx S_{nn''' \lambda_2 \lambda'_1}(t) S_{n'' n', \lambda_1 \lambda'_2}(t) + S_{nn' \lambda_2 \lambda'_2}(t) S_{n'' n''' \lambda_1 \lambda'_1}(t). \end{aligned} \quad (37)$$

With these approximations we finally get

$$\begin{aligned} m_{nn', \lambda \lambda'}(t) \approx & \sum_{n'', n'''} \sum'_{\lambda_1, \lambda'_1} \sum_{\lambda_2, \lambda'_2} V_{nn'; n'' n'''}^{\lambda \lambda'; \lambda_1 \lambda_2 \lambda'_1 \lambda'_2} . \\ & \cdot \left[ S_{nn''' \lambda_2 \lambda'_1}(t) S_{n'' n', \lambda_1 \lambda'_2}(t) + S_{nn' \lambda_2 \lambda'_2}(t) S_{n'' n''' \lambda_1 \lambda'_1}(t) \right] \end{aligned} \quad (38)$$

with the vertices

$$\begin{aligned} V_{nn'; n'' n'''}^{\lambda \lambda'; \lambda_1 \lambda_2 \lambda'_1 \lambda'_2} = & \left[ j(j+1) j'(j'+1) \right]^{-1} \sum_{\lambda_3 \lambda'_3} \sum_{\alpha, \alpha'} L_{\lambda \lambda_3}^\alpha L_{\lambda' \lambda'_3}^{\alpha'} \\ & \times b_{\lambda_1 \lambda_2 \lambda_3}^* b_{\lambda'_1 \lambda'_2 \lambda'_3} \langle l_n^\alpha | T_{nn''}^- [i \mathcal{L}_-^{(nn'' n''' n')}]^{-1} T_{n''' n'}^- | l_n^{\alpha'} \rangle. \end{aligned} \quad (39)$$

Equations (27)–(30), (38) and (39) are the closed set of MCT equations. Their mathematical structure is identical to original MCT in that sense that Eq. (27) is identical and that the memory kernel is bilinear in the t-dependent correlators. The main difference lies in the vertices. In the original version of MCT they depend on the static correlators, only. The modified version, Eq. (39), involves the calculation of the time dependent torque-torque correlator for two ( $n = n' \neq n'' = n'''$ ), three ( $n \neq n' \neq n'' = n'''$ ) and four rods (*all* $n, n', n'', n'''$  are different from each other). Note that

$$\begin{aligned} & \langle l_n^\alpha | T_{nn''}^- [i \mathcal{L}_-^{(nn'' n''' n')}]^{-1} T_{n''' n'}^- | l_n^{\alpha'} \rangle \\ & = \int_0^\infty dt \langle l_n^\alpha | T_{nn''}^- \exp[-i \mathcal{L}_-^{(nn'' n''' n')} t] T_{n''' n'}^- | l_n^{\alpha'} \rangle \end{aligned} \quad (40)$$

where the time dependence is given by  $\mathcal{L}_-^{(nn''n'''n')}$ .

The long time behavior of  $\mathbf{S}(t)$  follows from

$$\mathbf{S}(t) + \int_0^t dt' \mathbf{m}(t-t') \dot{\mathbf{S}}(t') = 0 \quad (41)$$

from which one obtains the equation for the non-ergodicity parameters (not normalized)  $\mathbf{F} = \lim_{t \rightarrow \infty} \mathbf{S}(t)$ :

$$\mathbf{F}[\mathbf{S}(0) - \mathbf{F}]^{-1} = \mathcal{F}(\mathbf{F}) \quad , \quad \mathcal{F}(\mathbf{F}) = \lim_{\mathbf{t} \rightarrow \infty} \mathbf{m}(\mathbf{t}). \quad (42)$$

### 3 Discussion and Conclusions

We have studied systems for which the glassy dynamics and the glass transition is not the consequence of growing static correlations but of entanglement effects. Since the direct correlation functions vanish for these systems, the original version of MCT can not be used because the memory kernels responsible for the slowing down of relaxation vanish. In order to find an alternative microscopic approach we have chosen a system of infinitely hard rods fixed with their centers on a periodic lattice. As a first step we have investigated the hierarchy of equations for the reduced probability density obtained from the Smoluchowski equation for  $N$  rods. This hierarchy has been truncated at its second level. This level yields the time variation of the 2-rod density as functional of the 1-, 2- and 3-rod density. Now, assuming that the influence of a third rod on the other two rods can be approximated by an effective diffusion tensor we finally have obtained a self-consistency equation for that tensor. This equation yields a decrease of the rotational diffusion when  $l$  is increased and an arrest at a critical length  $l_c$  [18,19]. The slowing down of the rotational dynamics and the glass transition is driven by the growth of the time dependent torque-torque correlator (Eq. (5)) with increasing  $l$ . Although a glass transition is obtained, there are two short-comings. First, the transition is of type-A [4], i.e. continuous (cf. [11]). Second, a consequence of a type-A transition is that the exponent  $\gamma$  equals one. These two results are in disagreement with the numerical ones [11]. The continuity of the transition is the most striking drawback.

In order to remove these discrepancies between the simulational and theoretical results we have returned to the exact Mori-Zwanzig equation for the orientational correlators  $S_{nn',\lambda\lambda'}(t)$ . Since we already know that the vertices of the approximated memory kernel of original MCT are zero (due to the trivial statics) we have resorted to an alternative. This has led to the factorization of the correlation function of the fluctuating forces into two parts. The first factor is the correlation function of pair densities, as in original MCT. Following MCT, this first part is approximated again by products

$S_{n_1 n'_1, \lambda_1 \lambda'_1}(t) S_{n_2 n'_2, \lambda_2 \lambda'_2}(t)$ . The second factor, the vertex, is determined by the time-dependent torque-torque correlator of a system of only two, three and four rods. The crucial input are their time integrals from zero to infinity (cf. Eq. (40)) which are equal to their Laplace transform at frequency  $z = 0$ . These time integrals are a measure of the collisional events. With increasing  $l$  the probability for collisions will increase, leading to an increase of the “area” under the time-dependent torque-torque correlators (cf. Fig. 2) and in turn to an increase of the vertices. Accordingly, the “area” under the time dependent torque-torque correlators plays the same role as the static correlators do for systems with excluded volume effects. Whether or not the  $l$ -dependence of the “area” under the torque-torque correlators is the same or different for Newtonian and Brownian dynamics is not obvious. It will be important to check this point, because two different critical lengths  $l_c$ , exponents  $\gamma$ , etc. would occur in case that this “area” has different  $l$ -dependence.

There is another point which has to be stressed. The existence of a discontinuous glass transition does not only require a bilinear memory kernel but also corresponding properties of the vertices. Sufficient but not necessary is their positive definiteness [4,21]. Whether a discontinuous transition really occurs has still to be shown for the modified MCT.

The same type of approximation of the memory kernel as presented in Sect. 2 can also be applied to particle systems with excluded volume interactions. Again, the growth of memory effects is ruled by the “area” under the time dependent force-force correlators of two, three and four particles. If this area becomes large enough at low temperatures a glass transition will occur, provided the vertices have corresponding properties. However, the  $q$ -dependence of the corresponding non-ergodicity parameters will not be in phase with the  $q$ -dependence of the corresponding static correlators as found from original MCT. This lack of the influence of static correlators on the glassy dynamics and glass order parameters (non-ergodicity parameters) for systems with dominant excluded volume effects is a drawback of the present mode coupling approximation. In this sense both MCT, the conventional and the present one, are complementary to each other. The conventional theory yields a satisfactory description of glassy dynamics and structural arrest for systems with strong excluded volume interactions and the modified theory may be suitable in cases where entanglement is the dominant mechanism. The latter does not require infinitely thin objects, like infinitely thin rods on a lattice or threads in a liquid. Even if the thickness is finite but small the static correlations may be still too weak to drive a glass transition in the framework of conventional MCT, but the present version of MCT may still yield a transition.

Whether the modified MCT also yields a satisfactory agreement with e.g. the MD-results [11] as original MCT does for numerous systems has still to be checked. If this will be true, it would be desirable to unify both mode coupling approximations in such a way that the resulting vertices contain

both, excluded volume and entanglement effects. This is a great challenge which will require a deeper insight into the quality and validity of the mode coupling approximations.

### Acknowledgements

The main new results of this work (Sect. refsec2) were obtained during a two months stay with W. Götze at the Physics Department of the Technical University of Munich. I am grateful for the hospitality. I also would like to thank W. Götze for stimulating discussions which have led to the modified version of MCT and for valuable comments on the manuscript. The fruitful collaboration with G. Szamel which has led to our first microscopic approach presented in Section 1, is acknowledged as well.

### References

1. J. Jäckle, Rep. Prog. Phys. **49**, 171 (1995).
2. R. Schilling in “*Collective Dynamics of Nonlinear and Disordered Systems*” eds. G. Radons, W. Just and P. Häussler, (Springer-Verlag, Berlin 2004); cond-mat 0305565.
3. U. Bengtzelius, W. Götze and A. Sjölander, J. Phys. C **17**, 5915 (1984).
4. W. Götze in “*Liquids, Freezing and the Glass Transition*”, eds. J.P. Hansen, D. Levesque and J. Zinn-Justin, North-Holland, Amsterdam 1991.
5. R. Schilling in “*Disorder Effects on Relaxational Processes*”, eds. R. Richert and A. Blumen, (Springer-Verlag, Berlin 1994).
6. H.Z. Cummins, J. Phys.: Condens. Matter **11**, A95 (1999).
7. W. Kob and H.C. Andersen, Transp. Theor. Stat. Phys. **24**, 1179 (1995); W. Kob, J. Phys.: Condens. Matter **11**, R85 (1999); W. Kob, Les Houches lecture notes 2002, cond-mat/0212344.
8. K. Binder, J. Baschnagel and W. Paul, Prog. Polym. Sci **28**, 115 (2003).
9. W. Götze and L. Sjörgen, Rep. Prog. Phys. **55**, 241 (1992); W. Götze, J. Phys.: Condens. Matter **11**, A1 (1999).
10. F. Ritort and P. Sollich, Adv. Phys. **52**, 219 (2003).
11. C. Renner, H. Löwen and J.L. Barrat, Phys. Rev. E **52**, 5091 (1995).
12. S.P. Obukhov, D. Kobsev, D. Perchak and M. Rubinstein, J. Physique I **7**, 563 (1997).
13. K.S. Schweizer, J. Chem. Phys. **91**, 5802, 58022 (1989).
14. V.G. Rostiashvili, Sov. Phys. JETP **70**, 563 (1990).
15. M. Fuchs and K.S. Schweizer, J. Chem. Phys. **106**, 347 (1997).
16. S.-H. Chong and M. Fuchs, Phys. Rev. Lett. **88**, 185702 (2002).
17. K. Miyazaki and A. Yethiraj, J. Chem. Phys. **117**, 10448 (2002).
18. R. Schilling and G. Szamel, Europhys. Lett. **61**, 207 (2003).
19. R. Schilling and G. Szamel, J. Phys.: Condens. Matter **15**, 967 (2003).
20. T. Aspelmeier, M. Huthmann and A. Zippelius in *Granular Matter*, Lecture Notes in Physics, eds. S. Ludwig and T. Pöschel, (Springer-Verlag, Berlin 2000).
21. T. Franosch and T. Voigtmann, J. Stat. Phys. **109**, 237 (2002).

## Part VIII

### Spin Systems

# Hydrogen and Magnetism in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$

Sebastian T. B. Goennenwein<sup>1,2</sup>, Thomas A. Wassner<sup>1</sup>, Hans Huebl<sup>1</sup>, Achim Koeder<sup>3</sup>, Wladimir Schoch<sup>3</sup>, Andreas Waag<sup>4</sup>, Jan B. Philipp<sup>5</sup>, Matthias Opel<sup>5</sup>, Rudolf Gross<sup>5</sup>, Martin Stutzmann<sup>1</sup>, and Martin S. Brandt<sup>1</sup>

<sup>1</sup> Walter Schottky Institut, Technische Universität München  
Am Coulombwall 3, 85748 Garching, Germany

<sup>2</sup> Department of Nanoscience, Delft University of Technology  
Lorentzweg 1, 2628 CJ Delft, The Netherlands

<sup>3</sup> Abteilung Halbleiterphysik, Universität Ulm  
89069 Ulm, Germany

<sup>4</sup> Institut für Halbleitertechnik  
Hans-Sommer-Straße 66, 38106 Braunschweig, Germany

<sup>5</sup> Walther-Meissner-Institut, Bayerische Akademie der Wissenschaften  
Walther-Meissner-Str. 8, D-85748 Garching, Germany

**Abstract.** Our recent studies of the influence of hydrogenation on the ferromagnetic semiconductor  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  are reviewed. We find that upon exposure to a remote DC hydrogen plasma, the hole density  $p$  in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  thin films with  $x = 0.037$  and  $x = 0.051$  can be reduced by several orders of magnitude, while the density of Mn magnetic moments is not significantly affected by the plasma treatment. The ferromagnetism clearly present in the as-grown samples vanishes after hydrogen incorporation. We analyze the effect of the hydrogenation in detail with the help of secondary ion mass spectroscopy, electronic transport, Hall, DC magnetization, and Fourier-transform infra-red absorption experiments. All results indicate that the Mn acceptors are electrically passivated by the formation of Mn-As-H complexes, resulting in the loss of long-range magnetic ordering.

Dilute magnetic semiconductors (DMS) are a very interesting class of materials, because they bring together the long-range magnetic order characteristic of ferromagnets and the versatile properties of semiconductors. Therefore, DMS are considered as an important building block for semiconductor-based magneto-electronic devices. Moreover, exploiting the well-established techniques of semiconductor technology to manipulate the semiconductor properties of DMS often also allows to control the magnetic ordering. Several such approaches have been demonstrated in different DMS, including the use of photo-generated charge carriers [1], the field effect [2], annealing [3], and co-doping [4].

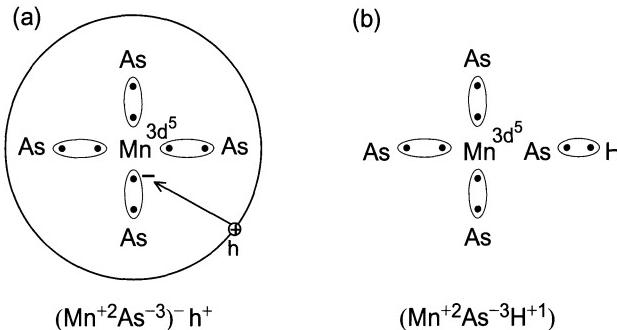
$\text{Ga}_{1-x}\text{Mn}_x\text{As}$ , on which we focus here, is currently the prototype III-V dilute magnetic semiconductor. For Mn contents  $x \geq 0.01$ ,  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  becomes ferromagnetic, with Curie temperatures  $T_C \leq 170$  K [5,6]. The introduction of Mn into GaAs simultaneously gives rise to (i) large localized magnetic moments and (ii) delocalized holes. The former are caused by the electrons in the half-filled  $3d^5$ -shell of Mn, which couple to a high-spin

$S = 5/2$  state according to Hund's rules. The itinerant holes are due to the fact that the two-valent Mn on a substitutional Ga site is an acceptor in GaAs. With the properties (i) and (ii), the two ingredients for itinerant ferromagnetism are present in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ . Indeed, it now is widely accepted that the ferromagnetic exchange between the Mn moments in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  is mediated by delocalized holes [6,7]. Therefore, a variation of either the density of Mn moments  $N_{\text{Mn}} \approx x$  or the hole density  $p$  should result in a change of the ferromagnetic properties of  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ , and thereby allow to critically test the theoretical models developed or to adjust the magnetic properties of the material for a specific application. However, an independent variation of  $N_{\text{Mn}}$  and  $p$  is not as straightforward as it might seem, as both are determined by the Mn content  $x$ , so that a post-growth control of the hole density in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  allowing to change  $p$  by orders of magnitude has proven elusive.

It is well established that hydrogen incorporation can result in a passivation or compensation of donors or acceptors. In particular, it has been demonstrated that the group-II acceptors Be, Mg and Zn in GaAs can be passivated via hydrogenation [8,9,10,11,12]. In this paper, we summarize our recent investigations which show that a hydrogen passivation of the Mn acceptor in GaAs is also possible [13], allowing to change the hole density  $p$  in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  thin films by several orders of magnitude, while the density of localized Mn magnetic moments is not significantly affected by the hydrogen incorporation. As a consequence, the ferromagnetism present in the as-grown samples gives way to paramagnetism after the hydrogenation – as expected upon the removal of the itinerant holes mediating the ferromagnetic exchange interaction. The experiments discussed here are focused on the two extreme situations (as-grown vs. completely passivated  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  films), but our results also show that a partial hydrogenation and thus a “hydrogen-control” of ferromagnetism is possible.

## 1 Sample Preparation and Hydrogenation

The  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  thin films were grown by molecular beam epitaxy on semi-insulating, (001)-oriented GaAs substrates. A nominally undoped GaAs buffer layer was first deposited at a substrate temperature  $T_S = 585^\circ\text{C}$  (high temperature, HT) to initiate the growth. Then, the substrate temperature was lowered to  $T_S = 270^\circ\text{C}$  for the low-temperature (LT) growth of the  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  thin film. In some samples, an additional undoped LT GaAs buffer layer was deposited prior the  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  layer. In the following, we will focus on two samples: a 320 nm thick  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  layer with  $x = 0.037$  and a 330 nm thick  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  layer with  $x = 0.051$ , respectively. The Mn content  $x$  was determined from high-resolution X-ray diffraction measurements [6].



**Fig. 1.** (a) In a simple electron counting picture, Mn residing on a substitutional Ga site acts as an acceptor, because one electron from the valence band is required to saturate all bonds. The delocalized hole wave function is schematically represented as a large circle. Assuming complete ionization, Mn and As are in the oxidation states +2 and -3, respectively, and the positively charged hole  $\text{h}^+$  is required to obtain charge neutrality. (b) Upon hydrogenation, a Mn-As-H complex is formed. Because the H atom supplies an additional electron to the complex, Mn does not act as an acceptor any more, which results in a decrease of the density of free holes. The oxidation states of  $\text{Mn}^{+2}$  and  $\text{As}^{-3}$  do not change upon hydrogenation, the charge neutrality of the complex now achieved by the ionized  $\text{H}^+$  atom

A remote DC hydrogen or deuterium plasma operated at 0.3 mbar was used to hydrogenate (deuterate) the  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  thin films. During the plasma treatment, the samples were heated to 170°C. These parameters were chosen based on a passivation study of Zn acceptors in GaAs:Zn [11]. Up to 168 h of plasma exposure time were required for a complete hydrogenation of the  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  films. We tentatively attribute this effect to the trap-limited diffusion of hydrogen; this issue is currently being investigated in detail and will be discussed in a future publication.

## 2 Hydrogen and Group-II Acceptors in GaAs

The effect of hydrogenation on the group-II acceptors Be, Mg and Zn in GaAs has been investigated in some detail [8,9,10,11,12]. Hydrogen was found to form complexes with these acceptors, passivating them. Local vibrational mode spectroscopy has proven useful to investigate the microscopic structure of these complexes. Fourier-transform infra-red spectroscopy (FTIR) experiments have shown that the H atom is located in a bond-centered position in GaAs:Be [9,10]. In contrast, As-H stretching modes around  $2145 \text{ cm}^{-1}$  are observed in hydrogenated GaAs:Mg and GaAs:Zn at low temperatures, the mode position being rather independent of the specific acceptor studied. This has been taken as evidence that the hydrogen is in a back-bonded configuration in the corresponding complexes, as indicated schematically in

Fig. 1(b). Due to the similar ionic radii of Mn and Zn, H in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  is also expected to go into the back-bonded position.

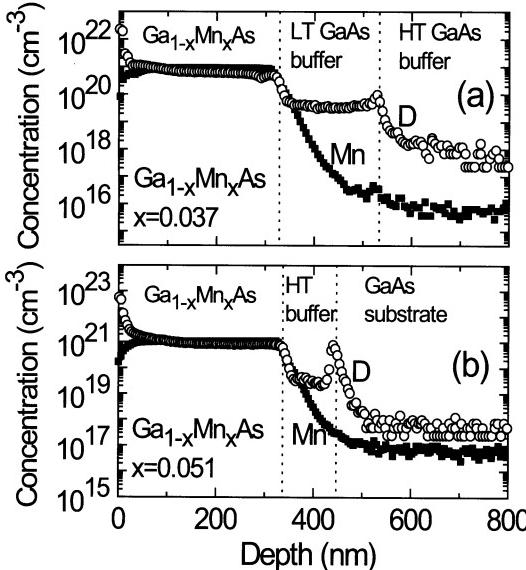
The effect of hydrogenation on the electronic properties of the acceptors can be understood in terms of an electron counting scheme, as illustrated in Fig. 1. As is the case for the other group-II acceptors, substitutional Mn on a Ga site in GaAs is short of one electron to saturate the four bonds to the neighboring As atoms. It therefore acts as an acceptor, taking an electron from the valence band and thereby introducing an itinerant hole into the crystal. Upon hydrogenation, the electron required to saturate the dangling As bond is provided by the H atom. Therefore, Mn in the Mn-As-H complex is expected to be passivated and should no longer act as an acceptor, as in the case of the other group-II acceptors. It is important to note that the Mn 3d-electrons are not affected by the hydrogenation according to this model.

### 3 Hydrogen in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ : Experimental Results

Assuming that the effect of the hydrogenation of Mn in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  is similar to that of the group-II acceptors Mg and Zn in GaAs summarized in the previous section, one would expect hydrogen to chemically bind in Mn-As-H complexes in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  after the hydrogenation. At the same time, the hydrogen treatment should result in a significantly reduced hole concentration. In the following, we compare these predictions with experiment. We start the discussion of our experiments with a quantitative analysis of the H incorporation.

#### 3.1 Secondary Ion Mass Spectroscopy

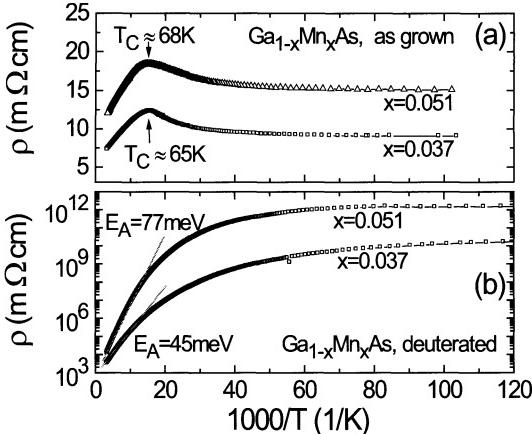
The concentration of hydrogen present in the  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  thin films after the plasma process was quantified by secondary ion mass spectroscopy (SIMS), using the hydrogen isotope deuterium which can be detected more sensitively due to its low natural abundance. The corresponding depth profiles are shown in Fig. 2, using  $\text{Cs}^+$  as primary ions and detecting Mn as positive and D as negative ions in a Cameca 5 spectrometer. The D concentration was determined quantitatively by comparison to implantation standards. For Mn, no such standards were available. We therefore have normalized the Mn concentrations given in Fig. 2 to the nominal Mn content determined from XRD. In both samples, the Mn and D concentrations are identical to within the experimental uncertainty of the SIMS measurements. Note also that the concentration of D in the GaAs buffer layers is significant. This indicates the presence of a high density of structural defects in these layers, which can trap deuterium. Similarly, the high concentration of D at the interfaces is indicative of a high density of deuterated defects there. However, although these defects may lead to compensation, their density is more than an order of magnitude smaller than the Mn concentration. Thus, the Mn doping indeed determines the electrical properties of the thin films.



**Fig. 2.** Secondary ion mass spectroscopy depth profiles of deuterated  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  thin films with (a)  $x = 0.037$  and (b)  $x = 0.051$  show that the concentration of incorporated deuterium is identical to the Mn concentration in the films to within experimental error

### 3.2 Electronic Transport

As expected for the high Mn doping levels, the as-grown  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  films are metallic. Figure 3(a) shows that their resistivity is nearly temperature independent. The small decrease of the resistivity for  $T \leq 70$  K in both samples is typically observed in ferromagnetic  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  films and has been attributed to a reduced spin-disorder scattering [5,14]. The resulting hump in the resistivity can be used as an estimate of the Curie temperature  $T_C$  [6]. In sharp contrast to this metallic behavior, the samples are semiconducting after deuteration (Fig. 3(b)). The resistivity at room temperature increases by about three orders of magnitude compared to the values prior to hydrogenation, and by up to 11 orders of magnitude at 4 K. Near room temperature, the resistivity is thermally activated with an activation energy  $E_A = (50 \pm 10)$  meV for  $x = 0.037$  and  $E_A = (80 \pm 10)$  meV for  $x = 0.051$ . These energies correspond to roughly half of the energy separation  $E_{\text{Mn}} = 113$  meV of the Mn acceptor ground state from the GaAs valence band edge [15], indicating that a small portion (about  $10^{-3}$ ) of the passivated Mn acceptors is still active and determines the resistivity, while compensation only plays a minor role. Room-temperature Hall measurements show that both deuterated films still are *p*-type, with hole densities  $p = 1.1 \times 10^{18} \text{ cm}^{-3}$  and  $p = 5.5 \times 10^{17} \text{ cm}^{-3}$  for  $x = 0.037$  and  $x = 0.051$ , respectively, supporting the conclusions drawn from the resistivity measurements. No indications for

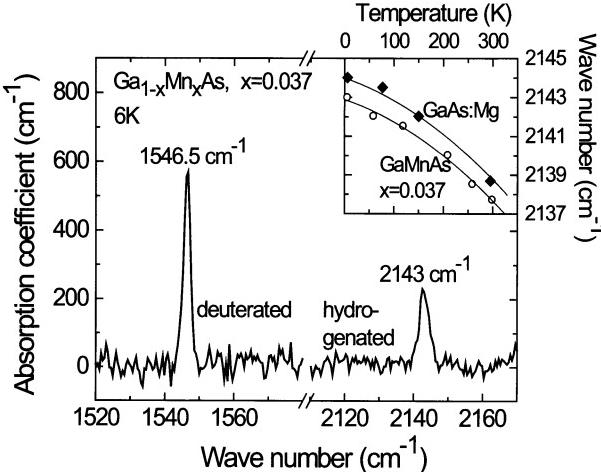


**Fig. 3.** Temperature-dependent dark resistivity of  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  thin films (a) as-grown and (b) after deuteration. The as-grown films are metallic, while semiconducting behavior is observed after the deuteration

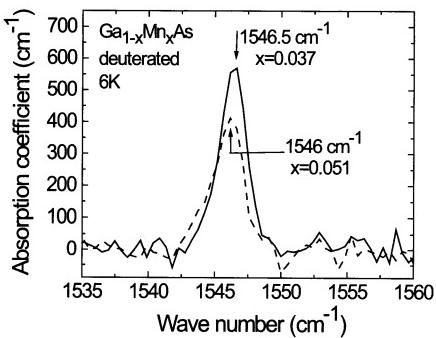
anomalous Hall contributions could be detected at 295 K. The deuteration therefore clearly results in a very efficient electronic passivation of the Mn acceptors.

### 3.3 Local Vibrational Mode Spectroscopy

To elucidate the microscopic state of H in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ , we have investigated the infrared absorption properties of both fully hydrogenated and fully deuterated  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  samples, using a Bruker IFS 113v FTIR spectrometer equipped with a Helium flow cryostat. Figure 4 shows the FTIR transmission spectra for samples with a Mn content  $x = 0.037$ . In the hydrogenated material, we find an absorption line at  $2143\text{ cm}^{-1}$  at  $T = 6\text{ K}$ . The line position changes with increasing temperature as shown in the inset of Fig. 4, its full width at half maximum increasing only marginally from  $3\text{ cm}^{-1}$  at  $T = 6\text{ K}$  to  $4\text{ cm}^{-1}$  at  $T = 295\text{ K}$ . In the deuterated  $\text{Ga}_{0.963}\text{Mn}_{0.037}\text{As}$  sample, the absorption line is shifted to  $1546.5\text{ cm}^{-1}$  at  $T = 6\text{ K}$ . The effect of the Mn content  $x$  on the vibrational mode appears to be rather small, as shown exemplarily for deuterated material in Fig. 5. Independent of our investigations, Bouanani-Rahbi and coworkers have performed similar studies of the vibrational properties of H in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  [16]. They find essentially the same As-H stretching mode at  $2140.5\text{ cm}^{-1}$  in  $\text{Ga}_{0.966}\text{Mn}_{0.034}\text{As}$ , the mode position shifting with temperature in close analogy to the results shown in Fig. 4 for our samples. The only difference to our results is that Bouanani-Rahbi et al. report a weak dependence of the mode position on the Mn concentration  $x$ , while we observe the modes to be independent of  $x$  to within experimental uncertainty (Fig. 5).

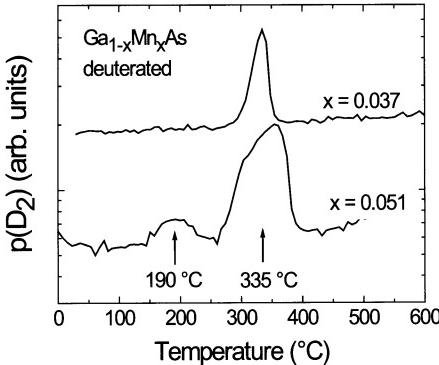


**Fig. 4.** Infrared absorption spectra of hydrogenated and deuterated  $\text{Ga}_{0.963}\text{Mn}_{0.037}\text{As}$ . The spectra were taken with a spectral resolution of  $0.5 \text{ cm}^{-1}$ . Inset: Temperature dependence of the As-H stretching mode in hydrogenated  $\text{Ga}_{0.963}\text{Mn}_{0.037}\text{As}$  (open circles, this work) and in hydrogenated  $\text{GaAs:Mg}$  (full squares, [12]). The lines are guides to the eye, revealing the close similarity between the modes of the two different complexes



**Fig. 5.** Infrared absorption spectra of deuterated  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  films with  $x = 0.037$  and  $x = 0.051$ . The influence of the Mn content  $x$  on the mode position is smaller than the spectral resolution

The As-H and As-D stretching modes which we observe in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  are very similar to those of the heavier group-II acceptors in  $\text{GaAs}$  [13]. The mode position is nearly independent of the specific acceptor considered, and even the temperature dependence of the As-H stretching mode in  $\text{Ga}_{0.963}\text{Mn}_{0.037}\text{As}$  and in  $\text{GaAs:Mg}$  are similar (Fig. 4, inset). Thus, our vibrational mode spectroscopy results suggest that hydrogen is in a back-bonded configuration also in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ . This is corroborated by X-ray diffraction, which shows that upon hydrogenation the lattice constant in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  increases much less than in  $\text{Si:B}$ , where H is incorporated on the bond center site [17]. However, further investigations still are necessary to fully clarify the microscopic structure of the Mn-As-H complexes.



**Fig. 6.** Effusion spectra (partial  $D_2$  pressure as a function of sample temperature) of deuterated  $Ga_{1-x}Mn_xAs$  films with  $x = 0.037$  and  $x = 0.051$

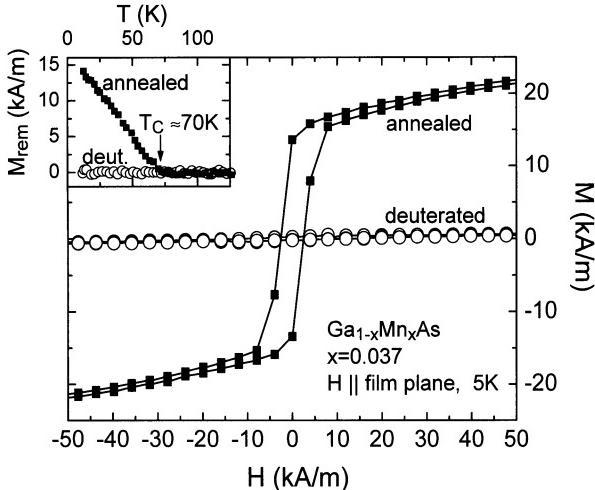
### 3.4 Thermal Effusion

To study the thermal stability of the Mn-As-H complexes, we have performed thermal effusion experiments. To this end, the partial pressure of  $D_2$  effusing from the sample was measured with a mass spectrometer while ramping up the sample temperature at a constant rate. Figure 6 shows the corresponding spectra for deuterated  $Ga_{1-x}Mn_xAs$  with  $x = 0.037$  and  $x = 0.051$ . In both samples, a pronounced effusion peak around  $335^\circ C$  is observed, with an additional, weaker peak around  $190^\circ C$  for  $x = 0.051$ . Similar experiments in deuterated  $GaAs:Zn$  showed two effusion peaks at  $200^\circ C$  and  $300^\circ C$ , which were attributed to D trapped at the surface and to dissociating D-acceptor-complexes, respectively [11]. Taking into account that as-grown  $Ga_{1-x}Mn_xAs$  films should not be exposed to temperatures significantly higher than their growth temperature  $T_S = 270^\circ C$  to avoid phase separation and cluster formation [6], the deuterated material can be considered as stable for temperatures up to the growth temperature.

## 4 Hydrogen and Magnetism in $Ga_{1-x}Mn_xAs$

Let us now discuss the magnetic properties of the films before and after the deuteration. The magnetization was measured using a superconducting quantum interference device (SQUID) magnetometer with a He cryostat. We will focus on two  $Ga_{0.963}Mn_{0.037}As$  samples, one annealed at  $170^\circ C$  for 168 h under a  $D_2$ -gas flow, the other deuterated with a remote plasma for 168 h at  $170^\circ C$ . Both samples thus have been exposed to the same thermal treatment and the same gas flow, while the DC glow discharge was on for the deuteration, and off for the annealing.

Figure 7 shows the magnetization  $M(H)$  as a function of the externally applied magnetic field  $H$  and the temperature dependence of the remanent magnetization  $M_{\text{rem}}$  of the two samples. The annealed film is clearly ferromagnetic, exhibiting hysteresis as well as a finite remanent magnetization up to the Curie temperature  $T_C \approx 70$  K. Although prolonged annealing at



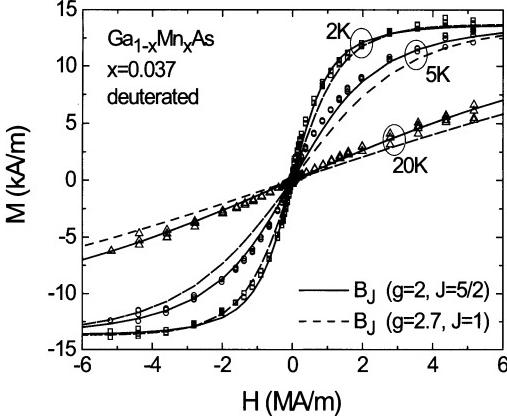
**Fig. 7.** Magnetization loop of  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  films with  $x = 0.037$  deuterated for 168 h at  $170^\circ\text{C}$  (open circles) and annealed under the same conditions, i.e. a constant flow of  $\text{D}_2$  gas, however without switching on the plasma discharge (full squares). Clearly, a change of the magnetization due to annealing can not account for the absence of ferromagnetism in the deuterated film. Inset: Temperature dependence of the remanent magnetization  $M_{\text{rem}}$  of the films, measured at  $H = 7.6$  kA/m

higher temperature can have a negative effect on the ferromagnetic properties of  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  [18], our results show that the above temperature treatment improves rather than destroys the ferromagnetism. This already has been observed by Edmonds et al. [19] in their low-temperature annealing experiments. The linear dependence of the remanent magnetization on temperature shown in the inset of Fig. 7 is often observed in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ , and has been attributed e.g. to a distribution of exchange constants [14,20]. We conclude that the magnetic properties of the annealed sample are those characteristic for state-of-the-art ferromagnetic  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ .

In contrast, no evidence for ferromagnetic behavior can be detected in the deuterated sample. As shown in Fig. 8, it has a Brillouin-type magnetization

$$M(H) = M_{\text{sat}} B_J(g\mu_B\mu_0 H J/(k_B T)) , \quad (1)$$

with the saturation magnetization  $M_{\text{sat}}$ , the Brillouin function  $B_J$ , the Bohr magneton  $\mu_B$ , the vacuum permeability  $\mu_0$ , the Boltzmann constant  $k_B$ , the  $g$ -value  $g$ , and the angular momentum quantum number  $J$ . According to the electron counting model in Fig. 1, one would expect  $g = 2$  and  $J = 5/2$  for isolated  $\text{Mn}^{+2}$  with a half-filled  $3d^5$  electron shell. Indeed, assuming these values in Eq. (1), a good agreement with experiment is obtained for all temperatures (Fig. 8, full lines). For comparison, we have also included the  $M(H)$ -dependencies expected for  $g = 2.7$ ,  $J = 1$  characteristic of Mn-hole complexes ( $3d^5 + \text{hole}$ ) in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  in Fig. 8 (dashed lines). Schneider et al. have



**Fig. 8.** Magnetization of the  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  film with  $x = 0.037$  deuterated for 168 h (cf. Fig. 7) at  $T = 2, 5$ , and  $20$  K (open symbols). The full and dashed lines correspond to a Brillouin function, Eq. (1), using  $g = 2, J = 5/2$  and  $g = 2.7, J = 1$ , respectively. A satisfactory fit is only obtained for  $g = 2, J = 5/2$ , corresponding to isolated  $\text{Mn}^{2+}$  in the  $3\text{d}^5$  configuration

found that these complexes are dominant in GaAs lightly doped with Mn [21]. The fact that the dashed lines do not describe the experimentally observed  $M(H)$ -data well shows that such Mn-hole complexes do not dominate the magnetic properties of hydrogenated  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ , in agreement with the conductivity and Hall results.

However, to obtain quantitative agreement between experiment and Eq. (1) in Fig. 8, one must assume  $M_{\text{sat}} = M_{\text{sat,deut}} \approx 15$  kA/m. While this value is somewhat smaller than the saturation magnetization  $M_{\text{sat,asgr}} \approx 20$  kA/m of the as-grown film, it still unambiguously shows that the localized magnetic moments are not significantly affected by the deuteration process. Nevertheless, both  $M_{\text{sat,asgr}}$  and  $M_{\text{sat,deut}}$  are more than a factor of 2 smaller than the theoretical value expected from  $M_{\text{sat,theo}} = xN_0g\mu_BJ \approx 38$  kA/m for  $x = 0.037$ , with the density of cation sites  $N_0 = 2.21 \times 10^{22}$  cm<sup>-3</sup> in GaAs and again taking  $g = 2, J = 5/2$ . Both the as-grown and the deuterated film thus exhibit a strong deficit in saturation magnetization. Such an effect has been repeatedly reported in ferromagnetic  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  [14,18,22], and has been attributed to e.g. Mn-Mn pairs which couple antiferromagnetically and therefore do not contribute to the net saturation magnetization [23,24]. On the other hand, the saturation magnetization  $M_{\text{sat,anneal}} \approx 36$  kA/m of the  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  film annealed at 170°C for a long period of time is large, nearly matching  $M_{\text{sat,theo}}$ . This might indicate that the low-temperature anneal leads to a substantial decrease of the density of antiferromagnetically coupled Mn-Mn pairs. In contrast, this effect appears to be completely suppressed during the deuteration, possibly because of the formation of deuterium-complexes with different diffusion properties. More detailed ex-

periments will be required to clarify this issue, in particular also addressing the effect of deuteration on films which already have been annealed and thus have  $M_{\text{sat,anneal}} \approx M_{\text{sat,theo}}$ .

Taken together, the magnetization experiments clearly demonstrate that the deuteration completely quenches the ferromagnetism present in the as-grown and annealed films. At the same time, the localized Mn magnetic moments are not significantly affected by the deuteration. Thus, the absence of ferromagnetism in the deuterated sample can not be due to an annealing effect, but must be attributed to the incorporation of deuterium and the corresponding reduction of the hole density.

## 5 Summary and Outlook

We have investigated the effect of a low-temperature hydrogen plasma treatment on the electronic, vibronic and magnetic properties of  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  thin films. Similar to the group-II acceptors Mg and Zn in GaAs, the hydrogenation results in the formation of Mn-As-H complexes and very effectively passivates the Mn acceptors. The localized Mn magnetic moments are not affected by the hydrogen incorporation. However, since the ferromagnetic exchange interaction in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  is mediated by delocalized holes, a control of the hole density via the hydrogenation allows a “hydrogen control” of ferromagnetism.

In this paper, we have only considered the two limiting cases, namely as-grown and fully hydrogenated samples, to demonstrate the viability of the method. However, a partial hydrogenation is also possible. Such a controlled hydrogen passivation should allow to reduce the hole density  $p$  by a well defined amount and induce controlled changes in the electronic and magnetic properties. Since it has been suggested that the magnetic anisotropy strongly depends on  $p$  [7], a patterned hydrogenation or hydrogen implantation might allow to vertically or laterally pattern the magnetic properties of  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  films, in terms of the Curie temperature or of the magnetic anisotropy. Moreover, such a local hydrogenation could make it possible to realize ferromagnet/semiconductor heterostructures in one single  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  film. This shows that the incorporation of hydrogen into ferromagnetic semiconductors opens up fascinating new areas for the combination of magnetic and semiconducting materials.

## Acknowledgements

The authors acknowledge financial support by the Deutsche Forschungsgemeinschaft (SFB 348).

## References

1. S. Koshihara, A. Oiwa, M. Hirasawa, S. Katsumoto, Y. Iye, C. Urano, H. Takagi, G. Munekata, Phys. Rev. Lett. , **78**, 4617 (1997).
2. H. Ohno, D. Chiba, F. Matsukura, T. Omiya, E. Abe, T. Dietl, Y. Ohno, K. Ohtani, Nature **408**, 944 (2000).
3. T. Story, R. R. Galazka, R. B. Frankel, P. A. Wolff, Phys. Rev. Lett. , **56**, 777 (1986).
4. M. E. Overberg, B. P. Gila, G. T. Thaler, C. R. Abernathy, S. J. Pearson, N. A. Theodoropoulou, K. T. McCarthy, S. B. Arnason, A. F. Hebard, S. N. G. Chu, J. M. Zavada, Y. D. Park, J. Vac. Sci. Technol. B **20**, 969 (2002).
5. F. Matsukura, H. Ohno, A. Shen, Y. Sugawara, Phys. Rev. B **57**, R2037 (1998).
6. H. Ohno: in *Semiconductor Spintronics and Quantum Computation*, D. D. Awschalom, D. Loss, N. Samarth (Eds.) (Springer, Berlin, 2002), p. 1.
7. T. Dietl, H. Ohno, F. Matsukura, Phys. Rev. B **63**, 195205 (2001).
8. B. Pajot, A. Jalil, J. Chevallier, R. Azoulay, Semicond. Sci. Technol. **2**, 305 (1987).
9. M. Stavola, S. J. Pearson, J. Lopata, C. R. Abernathy, K. Bergman, Phys. Rev. B **39**, 8051 (1989).
10. P. R. Briddon, R. Jones, Phys. Rev. Lett. **64**, 2535 (1990).
11. M. Stutzmann, J.-B. Chevrier, C. P. Herrero, A. Breitschwerdt, Appl. Phys. A **53**, 47 (1991).
12. J. Chevallier, B. Clerjaud, B. Pajot, in *Hydrogen in Semiconductors*, J. I. Pankove, N. M. Johnson (Eds.), Semiconductors and Semimetals **34** (Academic Press, Boston, 1991) pp. 447.
13. S. T. B. Goennenwein, T. A. Wassner, H. Huebl, M. S. Brandt, J. B. Philipp, M. Opel, R. Gross, A. Koeder, W. Schoch, A. Waag, Appl. Phys. Lett. **92**, 227202 (2004).
14. A. Van Esch, L. Van Bockstal, J. De Boeck, G. Verbanck, A. S. van Steenbergen, P. J. Wellmann, B. Grietens, R. Bogaerts, F. Herlach, G. Borghs, Phys. Rev. B **56**, 13103 (1997).
15. R. A. Chapman, W. G. Hutchinson, Phys. Rev. Lett. **18**, 443 (1967).
16. R. Bouanani-Rahbi, B. Clerjaud, B. Theys, A. Lemaitre, F. Jomard, Physica B **340-342**, 284 (2003).
17. M. Stutzmann, J. Harsanyi, A. Breitschwerdt, C. P. Herrero, Appl. Phys. Lett. **52**, 1667 (1988).
18. S. J. Potashnik, K. C. Ku, R. Mahendirian, S. H. Chun, R. F. Wang, N. Samarth, P. Schiffer, Phys. Rev. B **66**, 012408 (2002).
19. K. W. Edmonds, K. Y. Wang, R. P. Campion, A. C. Neumann, N. R. S. Farley, B. L. Gallagher, C. T. Foxon, Appl. Phys. Lett. **81**, 499 (2002).
20. M. P. Kennett, Mona Berciu, R. N. Bhatt, Phys. Rev. B **65**, 115308 (2002).
21. J. Schneider, U. Kaufmann, W. Wilkening, M. Baeumler, F. Köhl, Phys. Rev. Lett. **59**, 240 (1987).
22. P. A. Korzhavyi, I. A. Abrikosov, E. A. Smirnova, L. Bergqvist, P. Mohn, R. Mathieu, P. Svedlindh, J. Sadowski, E. I. Isaev, Yu. Kh. Vekilov, O. Eriksson, Phys. Rev. Lett. **88**, 187202 (2002).

23. K. M. Yu, W. Walukiewicz, T. Wojtowicz, I. Kuryliszyn, X. Liu, Y. Sasaki, J. K. Furdyna, Phys. Rev. B **65**, 201303 (2002).
24. J. Blinowski, P. Kacman, Phys. Rev. B **67**, 121204 (2003).

# Resonant Soft Energy X-Ray Diffraction of Charge, Spin and Orbital Ordering

Peter D. Hatton<sup>1</sup>, Stuart B. Wilkins<sup>2,4</sup>, Thomas A. W. Beale<sup>1</sup>,  
Tarnjit Johal<sup>3</sup>, Peter Bencok<sup>4</sup>, D. Prabhakaran<sup>5</sup>, and  
Andrew T. Boothroyd<sup>5</sup>

<sup>1</sup> Department of Physics, University of Durham

Rochester Buildings, South Road, Durham, DH1 3LE, UK.

<sup>2</sup> European Commission, Joint Research Center, Institute for Transuranium  
Elements

Hermann von Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

<sup>3</sup> Daresbury Laboratory, Warrington  
Cheshire, WA4 4AD, UK

<sup>4</sup> European Synchrotron Radiation Facility, Boîte Postal 220, 38043 Grenoble  
Cedex, France

<sup>5</sup> Department of Physics, University of Oxford, Clarendon Laboratory  
Parks Road, Oxford, OX1 3PU, UK

**Abstract.** Soft X-ray resonant diffraction is a new technique pioneered by our group. Huge resonant enhancements of charge and magnetic scattering can be obtained at the  $L$ -edges of  $3d$  transition metal oxides. We review the technique of soft X-ray diffraction as well as giving recent examples of charge, spin and orbital ordering from bulk single crystals. We have studied the low temperature phase of  $\text{La}_{0.5}\text{Sr}_{1.5}\text{MnO}_4$ . Previous claims of orbital ordering in such materials have relied on observations at the  $K$  edge. These claims have attracted a barrage of controversy from theoretical studies. Instead we have employed resonant soft X-ray scattering at the manganese  $L_{III}$  and  $L_{II}$  edges which provides a direct measurement of the orbital ordering. Energy scans at constant wavevector have been compared to theoretical predictions and show that at all temperatures there are two separate contributions to the observed scattering, direct Goodenough orbital ordering and strong cooperative Jahn-Teller distortions of the  $\text{Mn}^{3+}$  ions.

## 1 Introduction

The use of soft X-ray energies in single crystal diffraction experiments has received little attention due to the perception that the absorption by the sample is far too high. Certainly X-rays at such energies are extremely inconvenient to work with. Firstly, air absorption at such energies is severe; the experiment (sample, diffractometer etc) must be placed in vacuum. Secondly, soft X-rays have prohibitively long wavelengths, which means that the Ewald sphere is very small, and only a very small volume of reciprocal space around the origin is accessible. As a result, there have been no reported experiments of diffraction from single crystals using soft X-rays until very recently [1]. We

chose to study the bilayer manganite  $\text{La}_{2-2x}\text{Sr}_{1+2x}\text{Mn}_2\text{O}_7$  [2] which has unit cell parameters of  $3.87 \times 3.87 \times 20.1 \text{ \AA}$ . It displays colossal magnetoresistance and displays charge ordering at approximately 220 K with anti-ferromagnetic spin ordering occurring below  $\sim 180$  K. The  $c$ -axis is sufficiently long to allow us to be able to access the (002) Bragg reflection, even at soft X-ray energies.

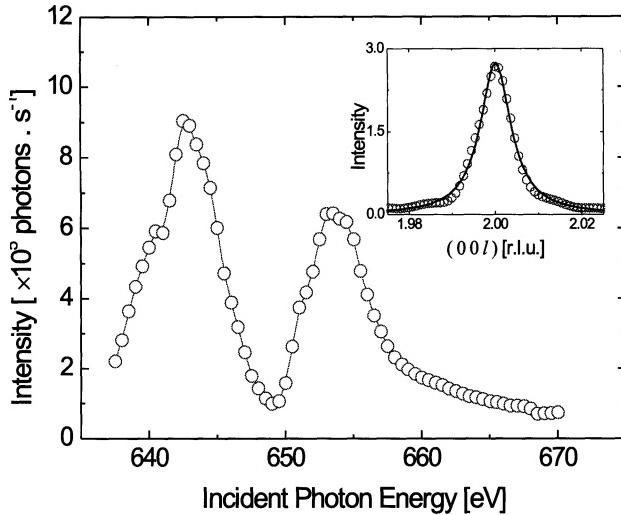
## 2 Charge Scattering

These experiments were conducted on station 5U.1 at the Synchrotron Radiation Source (SRS) at Daresbury Laboratory. The beamline produces a tunable monochromatic beam in the energy range 60–1500 eV. The two-circle diffractometer [3] is in a high vacuum vessel to eliminate air absorption, and a variable temperature stage using liquid nitrogen is located on the sample ( $\theta$ ) axis. The scattered photon flux was estimated from calibration of the drain current of the photodiode detector.

The crystal of  $\text{La}_{2-2x}\text{Sr}_{1+2x}\text{Mn}_2\text{O}_7$  with  $x = 0.475$  was grown by the floating-zone method. The sample was aligned in the diffractometer with the (001) and (110) defining the diffraction plane. Initially experiments were carried out in the non-resonant condition at a photon energy of 900 eV. Performing such a calculation to find the depth penetration of the incident beam leads to a correlation length of 4650 Å. This demonstrates that the X-rays are penetrating through many unit cells of the single crystal.

Next we recorded the (002) Bragg peak intensity as a function of the incident X-ray energy. Close to absorption edges the real and imaginary scattering factors are radically altered, giving increased absorption, and also increased scattering. Figure 1 shows the intensity of the (002) Bragg reflection as a function of energy close to the  $L_{III}$  and  $L_{II}$  absorption edges. Dramatic enhancement of the diffraction intensity is observed at  $\sim 640$  and  $\sim 652$  eV. These enhancements are so large that potentially extremely weak charge scattering, unobservable at off-resonant energies, could be resonantly enhanced to become easily observable via the resonant enhancement effect.

These charge scattering measurements already extend the range of possibilities of X-ray diffraction in the study of crystal structures, but the really exciting prospect is the use of such resonances in the study of magnetic structures. The use of X-ray scattering to probe the magnetic structure of crystalline solids started over 25 years ago with the study of NiO by de Bergevin and Brunel [4]. However, such scattering is an extremely weak phenomenon, typically  $10^8$  times weaker than charge reflections, with the result that most magnetic structure determination has employed neutron diffraction. Huge increases in the magnetic cross-section of X-ray scattering have been obtained by tuning the X-ray energy near to the absorption edge of the material [5,6,7,8]. These resonant enhancements can be so large that resonant magnetic scattering can be employed in the study of thin films and even surface layers.

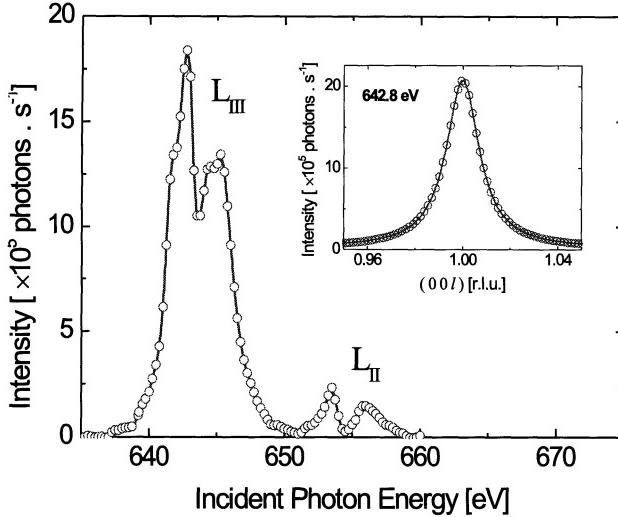


**Fig. 1.** An energy scan at constant wavevector through the (002) Bragg reflection at 300 K. The inset shows a scan in the  $\theta$ - $2\theta$  direction

Until now it has not been possible to exploit fully resonant magnetic diffraction to study 3d transition metal compounds, the largest class of magnetic materials in technological applications and fundamental research. Scattering at the  $K$  edge resonances of 3d ions has been observed [9,10], but the enhancement is small (factor 3–5) because the transition probed is the  $1s \leftrightarrow 4p$  while the spins responsible for the magnetic ordering belong to the 3d band. To study magnetism in transition metal compounds directly one needs to access the  $L_{III}$  and  $L_{II}$  edges, which exist in the region 500–900 eV.  $\text{La}_{2-2x}\text{Sr}_{1+2x}\text{Mn}_2\text{O}_7$  ( $x = 0.475$ ) undergoes a transition into a charge and orbitally ordered state at 220 K followed by a transition into a A type anti-ferromagnetic (AFM) structure at 180 K. Therefore below 180 K a magnetic diffraction peak due to the AFM structure should occur at the (001) position, even though the (001) Bragg reflection is systematically absent because of the body centering. Similar peaks (e.g., the (102)) have been previously observed by neutron single crystal studies [11].

### 3 Magnetic Diffraction

The sample was cooled down to 84 K and the incident x-ray energy tuned to the  $L_I$  absorption edge (769 eV). No resonance of the AFM (001) at this energy was observed. The incident photon energy was set to 639 eV corresponding to the  $L_{III}$  absorption edge. Such an edge causes a dipole transition from the  $2p^{3/2}$  to 3d levels directly probing the split electronic structure due to the magnetism. Appreciable intensity of the (001) was found at this wavevector, and a  $\theta$ - $2\theta$  scan is shown as the inset of Fig. 2. A scan of incident photon



**Fig. 2.** An energy scan at fixed wavevector through the (001) antiferromagnetic ordering reflection at 83 K. The inset shows a scan through the same peak in the (00l) direction

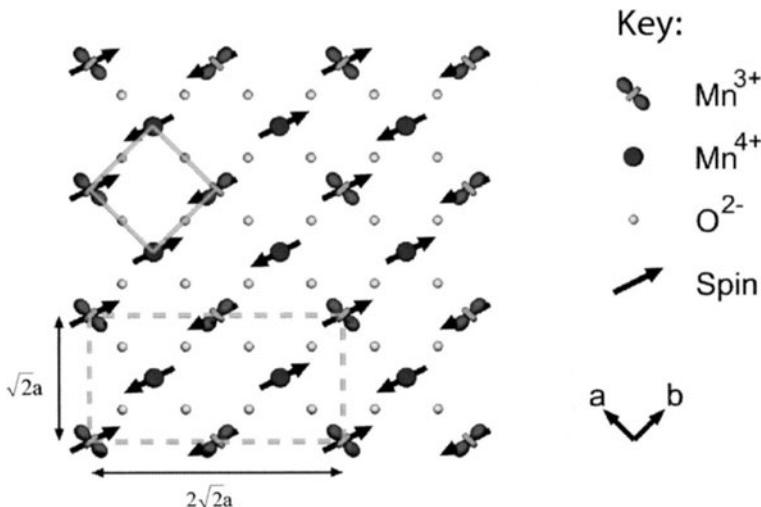
energy versus the integrated intensity of the (001) diffraction was measured. The result shown in Fig. 2, indicates two noticeable peaks showing an extremely large resonant enhancement at the  $L_{III}$  edge. In addition magnetic diffraction was also observed at the  $L_{II}$  edge. Away from the  $L_{III}$  and  $L_{II}$  resonances the magnetic peak was undetectable due to its very low intensity. Such a large resonant enhancement has been predicted quantitatively [12]. The different spectral shape observed for the (001) as compared to the (002) is most probably due to the electronic structure of the 3d band and its splitting due to Jahn-Teller distortions and the structure factor of the (001) being only sensitive to manganese ions. To confirm the origin of the reflection the sample was warmed to 300 K, past the Néel point at 180 K and the energy scan was repeated. No resonance was found at high temperature.

#### 4 Magnetic and Orbital Ordering

Orbital ordering, which involves correlations between the spatial distribution of the outermost valence electrons, has long been considered as a vital ingredient in the structural and physical properties of strongly correlated electron systems such as transition metal oxides. The competition and cooperation between the charge, orbital, and spin degrees of freedom of the electron manifests itself in unusual properties such as high temperature superconductivity, colossal magnetoresistance and magnetostructural transitions [14]. In particular, charge-orbital ordering of half-doped manganites has attracted much attention and controversy [15,16,17,18,19,20]. In  $\text{La}_{0.5}\text{Sr}_{1.5}\text{MnO}_4$  the Mn sites

at room temperature are all crystallographically equivalent with an average valency of +3.5. This material displays a phase transition at  $\sim 240\text{ K}$  below which additional reflections at a wave vector of  $(\frac{1}{2}, \frac{1}{2}, 0)$  have been observed caused by charge disproportionation of the Mn ions into two inequivalent sites identified as  $\text{Mn}^{3+}$  and  $\text{Mn}^{4+}$  (Fig. 3) [21,22,23]. The strong Hund's rule coupling and the large cubic ( $O_h$ ) component of the crystal field implies that the  $\text{Mn}^{3+}3d^4$  site has three electrons in the  $t_{2g\uparrow}$  level and one electron in the twofold degenerate  $e_{g\uparrow}$  level. The degeneracy of the  $e_{g\uparrow}$  level can be lifted by cooperative Jahn-Teller distortions of the  $\text{MnO}_6$  octahedral reducing the symmetry to  $D_{4h}$  and separating the two degenerate levels into  $3d_{3z^2-r^2}$  and  $3d_{x^2-y^2}$  levels (Fig. 4). The presence of Jahn-Teller distortions leads to additional satellites observed at a wave vector of  $(\frac{1}{4}, \frac{1}{4}, 0)$ . Therefore cooperative Jahn-Teller distortions of the  $\text{MnO}_6$  octahedra could provide the origin of orbital ordering in the manganites due to occupation of the  $3d_{3z^2-r^2}$  orbitals along the elongation direction. An alternative description as to how the orbital ordering may occur was provided by Goodenough [24]. Further cooling below  $T_N \sim 110\text{ K}$  results in long range spin ordering with a wave vector of  $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$  and the formation of a complex antiferromagnetic ordering of both manganese sublattices. The arrangements of the spins in  $\text{La}_{0.5}\text{Sr}_{1.5}\text{MnO}_4$  is similar to that adopted by the half doped manganites – the so-called magnetic CE phase [25]. This structure was intuitively predicted by Goodenough [24] who also predicted that the spin arrangements would occur via orbital ordering in the higher temperature phase. Thus the alternative model of orbital ordering highlights orbital order correlations caused by antiferromagnetic spin ordering. Until very recently such questions remained unanswered as direct observation of orbital ordering remained elusive.

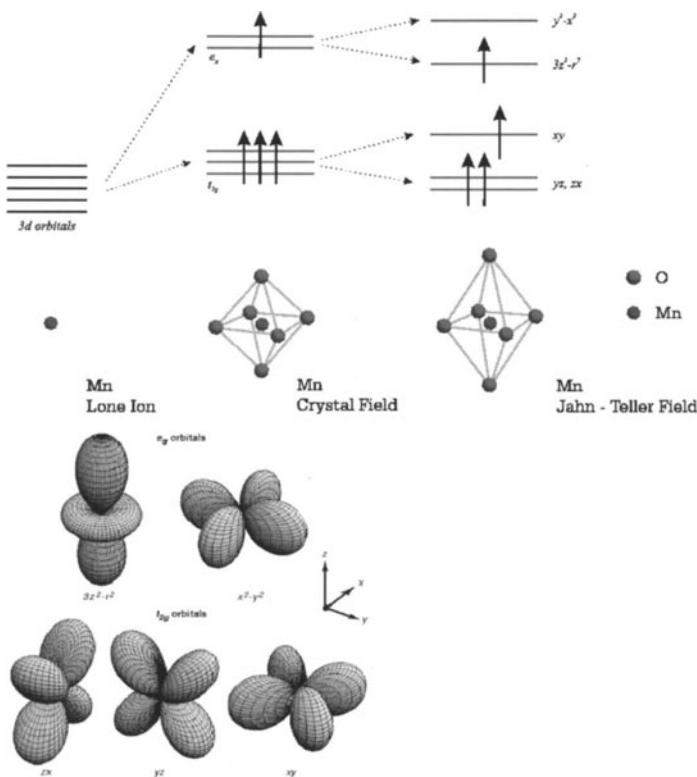
X-ray diffraction is generally sensitive to the isotropic electron density distribution. However by tuning the photon energy to an absorption edge enhanced sensitivity to the valence states is obtained [5,26]. Resonant X-ray Scattering (RXS) involves virtual excitations from core to valence states; photons are virtually absorbed by exciting core electrons to empty states, and subsequently reemitted when the excited electrons and their core hole recombine [27,6]. These processes lead to anomalous tensor components in the atomic scattering factors and hence to increased scattering and to scattering at positions forbidden by the crystallographic space group. Such scattering can be caused by long range ordering of magnetic moments, electronic orbital occupancy or aspherical valence electron clouds. And such scattering is element specific. The intensity of the superlattice reflections increase dramatically as the photon energy is tuned to an atomic absorption edge. RXS studies of  $\text{La}_{0.5}\text{Sr}_{1.5}\text{MnO}_4$  were first attempted at the manganese  $K$  edge. Resonant diffraction of an orbital ordering reflection displayed a striking increase near the absorption edge and an azimuthal dependence [23]. However the Mn  $K$  edge resonance involves virtual excitations from the  $1s$  to the  $4p$  band and is thus generally insensitive to orbital ordering of the  $3d$  levels. Theoret-



**Fig. 3.** Charge, orbital, and spin ordering in the  $\text{MnO}_2$  planes of  $\text{La}_{0.5}\text{Sr}_{1.5}\text{MnO}_4$ ; the arrows represent the magnetic moments. The  $I4/mmm$  unit cell is shown by the solid line ( $a = b = 3.864\text{\AA}$ ) and the OO unit cell by the dashed line

cal studies have proposed that the observed sensitivity is actually caused by Jahn-Teller distortions,  $4p$  band structure effects and  $3d$ - $4p$  Coulomb interactions [28,29,30,31,32,33,34]. In order to directly observe  $3d$  orbital ordering we need to access the Mn  $L$  edges. Resonant soft X-ray diffraction at the Mn  $L$  edges was first reported by Wilkins *et al.*, [1] and also the first direct observations of orbital ordering using soft X-rays in  $\text{La}_{0.5}\text{Sr}_{1.5}\text{MnO}_4$  [35]. Aided by published theoretical predictions [36] they reported that the shapes of the energy resonances at the Mn  $L_{III}$  and  $L_{II}$  edges showed that the orbital ordering was caused by a mixture of both cooperative Jahn-Teller distortions and direct Goodenough orbital correlations. These results, analysis and conclusions were later independently verified by another group [37].

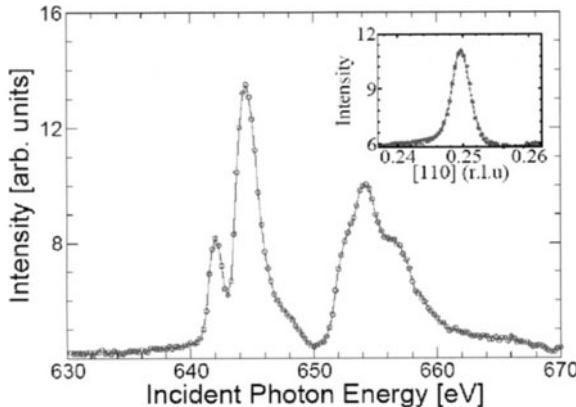
The experiments were performed at beamline 5U.1 at Daresbury and beamline ID08 at the European Synchrotron Radiation Facility. Single crystals of  $\text{La}_{0.5}\text{Sr}_{1.5}\text{MnO}_4$  with dimensions  $10 \times 3 \times 3\text{ mm}$  were grown at the University of Oxford using the floating zone method. They were cut with either [110] or [112] surface normal and polished with  $1\text{ }\mu\text{m}$  diamond paste to a flat shiny surface. The crystals were mounted in the ID08 five circle diffractometer operating at a base pressure of  $1 \times 10^{-10}\text{ mbar}$ , and equipped with a Si diode detector mounted behind variable rectangular aperture. The beamline produces  $\sim 100\%$  linearly polarized X-rays with an incident flux of  $10^{12}\text{ photons.s}^{-1}$  at  $\sim 650\text{ eV}$  with an energy resolution of  $350\text{ meV}$ . Sample cooling was achieved using a liquid He cryostat attached to the sample stage by copper braids resulting in a base temperature of  $63\text{ K}$ . Variable tem-



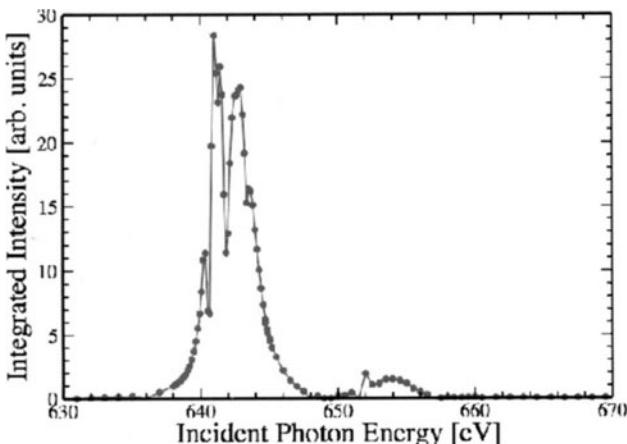
**Fig. 4.** Schematic diagram of the 3d band energy levels within a lone manganese ion (top), an octahedral crystal field (center), and a Jahn-Teller distorted tetragonal crystal field (bottom). The shapes of the 3d atomic orbitals are shown on the right, taken from <http://folk.uio.no/ravi/activity/ordering/orbitalordering.html>

perature scans of the magnetic and orbital reflections were undertaken at beamline 5U.1 at Daresbury Laboratory.

Figure 3 shows the generally accepted ground state displaying charge, spin and orbital ordering in the basal MnO<sub>2</sub> planes. The high temperature I4/mmm unit cell is shown by the solid line ( $a = b = 3.864 \text{ \AA}$  and  $c = 12.40 \text{ \AA}$ ). The occurrence of orbital order reflections with a wavevector of  $(\frac{1}{4}, \frac{1}{4}, 0)$  shows that the orbital ordering causes the unit cell to quadruple in the  $a - b$  plane. The new orbital ordered unit cell is shown by the dashed lines. The  $(\frac{1}{4}, \frac{1}{4}, 0)$  reflection was found to have a Lorentzian squared line shape. Figure 5 shows an energy scan at constant wavevector at 140 K through the  $(\frac{1}{4}, \frac{1}{4}, 0)$  peak through the Mn  $L_{III}$  and  $L_{II}$  edges at  $\sim 643$  and  $\sim 6564 \text{ eV}$ . Further confirmation that this scattering originates from anisotropic electron density effects is the result of an azimuth scan which shows the theoretically predicted behavior showing a  $\sin^2 \phi$  behavior. Figure 6 shows the energy scan



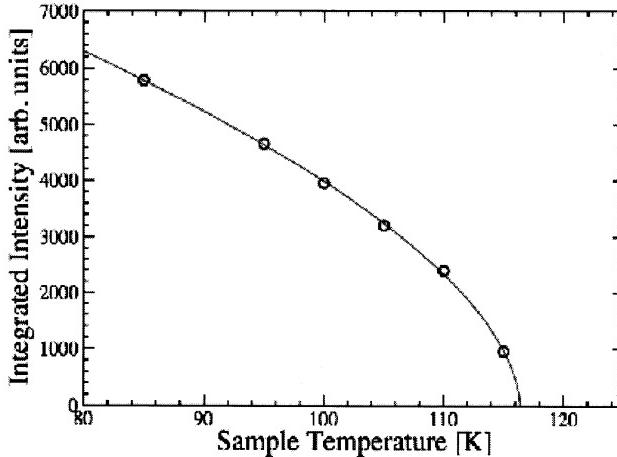
**Fig. 5.** Energy scan at fixed wavevector through the  $(\frac{1}{4}, \frac{1}{4}, 0)$  orbital order reflection at 140 K. Inset: a scan taken at  $E = 641$  eV through the orbital order reflection in the  $[110]$  (longitudinal) direction. By fitting the peak to a Lorentzian squared lineshape, the inverse correlation length is found to be  $2.6 \times 10^{-3} \text{ \AA}$



**Fig. 6.** Energy scan at fixed wavevector through the  $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$  magnetic diffraction peak at a temperature of 63 K

through the magnetic  $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$  reflection at 63 K. The energy scan at constant wavevector is considerably more intense than the orbital reflection and the spectral weight is far more concentrated at the Mn  $L_{III}$  edge.

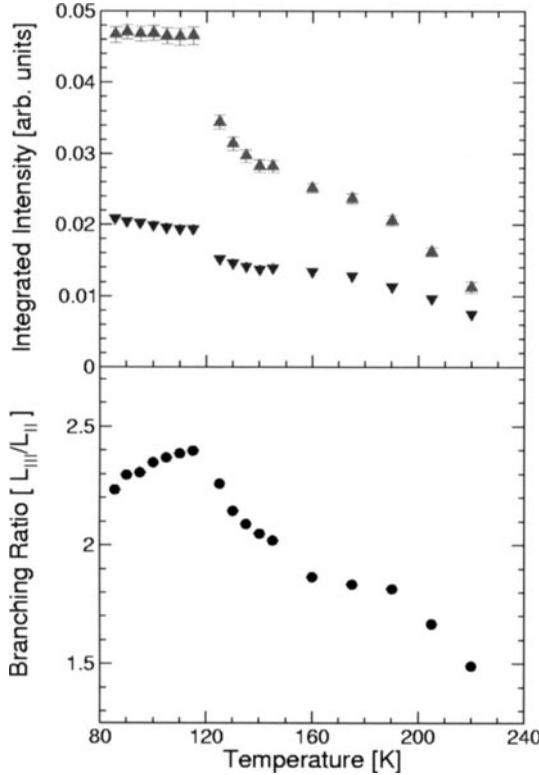
The integrated intensity of the magnetic  $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$  reflection at an energy of 642 eV is shown in Figure 7. This shows that the reflection is not observed, as expected, above  $T_N$ . Below  $T_N$  it increases in intensity and continues to increase in intensity as the temperature was lowered towards the base temperature. The solid line in Figure 7 is a fitted line displaying the order parameter. Figure 8 shows the integrated intensity of features at 643 and 653 eV in the



**Fig. 7.** The variation of the integrated intensity of the magnetic reflection  $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$  as a function of temperature below the Néel temperature

$(\frac{1}{4}, \frac{1}{4}, 0)$  orbital reflection as a function. As discussed in our earlier paper [35] and in [36] scattering at the Mn  $L_{III}$  edge ( $\sim 643$  eV) is primarily associated with cooperative Jahn-Teller distortions. Those at the Mn  $L_{II}$  ( $\sim 653$  eV) edge are primarily associated with direct Goodenough orbital ordering. As can be seen the features associated with the Jahn-Teller distortions are greatest at all temperatures below  $T_{OO}$ . Close to  $T_{OO}$  the features are of more nearly equal intensity. That means that at  $T_{OO}$  orbital order correlations caused by antiferromagnetic spin correlations are almost as important as the effects of the cooperative Jahn-Teller distortions in determining the long range orbital ordering. However upon further cooling the features at the Mn  $L_{III}$  edge grow with respect to those features at the  $L_{II}$  edge. Therefore although both Jahn-Teller distortions and antiferromagnetic orbital interactions increase as the temperature is lowered, those due to Jahn-Teller distortions increase at a greater rate. Of course the extent of the Jahn-Teller distortions would be expected to increase as the temperature is lowered. However although X-ray measurements show the Jahn-Teller distortions to have maximized by  $\sim 150$  K the effect of the Jahn-Teller distortions on the long-range orbital correlations continues to grow, both absolutely and with respect to the antiferromagnetic orbital correlations. At  $T_N$  ( $\sim 110$  K) an abrupt and unexpected increase in the intensity of the orbital order reflection is noted. There occurs a stepwise increase at both Mn  $L$  edges. Figure 8 shows the branching ratio – the ratio of the intensity of the  $L_{III}$  features over that of the  $L_{II}$  features. This ratio is always above 1.0 (Jahn-Teller features dominant) but maximizes at almost 2.5 at  $T_N$ , before decreasing somewhat.

The Goodenough model predicts a strong coupling between the orbital and magnetic correlations within the Mn<sup>3+</sup> sub-lattice. Magnetic susceptibility measurements give a much higher antiferromagnetic transition tem-



**Fig. 8.** The temperature dependence of the integrated intensity of the features in the energy spectrum (upper panel) at 643 eV (red triangles), and at 653 eV (blue inverted triangles). The lower panel shows the ratio of these two (the branching ratio ( $L_{III}/L_{II}$ ))

perature, concurrent with the charge and orbital ordering [36]. It therefore seems likely that in-plane antiferromagnetic order develops at 240 K ( $T_N(ab) = T_{CO} = T_{OO}$ ) consistent with the rod-like neutron scattering reported between  $T_N$  and  $T_{CO}$  [17]. This in-plane antiferromagnetic order then develops fully three dimensionally at  $T_N(c) = 110$  K. In this paper we have reported soft X-ray resonant diffraction on  $\text{La}_{0.5}\text{Sr}_{1.5}\text{MnO}_4$  at the Mn  $L_{III}$  and  $L_{II}$  edges for both the orbital order ( $\frac{1}{4}, \frac{1}{4}, 0$ ) and the magnetic order ( $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$ ) reflections. Our results allow the first direct comparison between orbital and magnetic correlations. Temperature dependant measurements of the intensity of different features in the energy spectra show large variations suggesting a very different temperature dependence of the two degrees of freedom and major changes at  $T_N$ . These results demonstrate the strong interaction between orbital and magnetic degrees of freedom, which we believe

to be a general property of many manganites and other  $3d$  transition metal compounds.

## 5 Conclusions

To conclude soft X-ray diffraction techniques offer a new method of resonantly increasing the scattering from very weak scattering, applicable to many  $3d$  transition metal compounds. By tuning to elemental absorption edges we can resonantly increase, by many orders of magnitude, the intensity of charge scattering. Indeed so large are the resonances that even magnetic scattering is easily observed. Finally resonant soft X-ray diffraction can also be used to directly observe orbital ordering. Indeed in the  $3d$  transition metal compounds resonant scattering at the  $L$  edges may be the only technique for directly observing orbital ordering.

## Acknowledgements

We are grateful to Daresbury Laboratory and to EPSRC and CLRC for financial support. SBW would like to thank the European Commission for support in the frame of TMR. PDH thanks the University of Durham for a Sir James Knott research fellowship and research leave.

## References

1. S.B. Wilkins, P.D. Hatton, M.D. Roper, D. Prabhakaran, A.T. Boothroyd, Phys. Rev. Lett. **90**, 187201 (2003).
2. Y. Moritomo, A. Asamitsu, H. Kuwahara, Y. Tokura, Nature **380**, 141 (1996).
3. M.D. Roper, G. van der Laan, H.A. Dürr, E. Dudzik, S.P. Collins, M.C. Miller, S.P. Thompson, Nucl. Instrum. Meth. Phys. Res. A **467**, 1101 (2001).
4. F. de Bergevin, M. Brunel, Acta. Cryst. A **37**, 314 (1981).
5. D. Gibbs, D.R. Harshman, E.D. Isaacs, D.B. McWhan, D. Mills, C. Vettier, Phys. Rev. Lett. **61**, 1241 (1988).
6. M. Blume, D. Gibbs, Phys. Rev. B **37**, 1779 (1988).
7. M. Blume, D. Gibbs, Phys. Rev. B **40**, 5218 (1989).
8. E.D. Isaacs, D.B. McWhan, C. Peters, G.E. Ice, D.P. Siddons, J.B. Hastings, C. Vettier, O. Vogt, Phys. Rev. Lett. **62**, 1671 (1989).
9. W. Neubeck, C. Vettier, K.-B. Lee, F. de Bergevin, Phys. Rev. B. **60**, R9912 (1999).
10. J.P. Hill, D.F. McMorrow, A.T. Boothroyd, A. Stunault, C. Vettier, L.E. Berman, M. von Zimmermann, Th. Wolf, Phys. Rev. B. **61**, 1251 (2000).
11. T. Chatterji, G.J. McIntyre, W. Caliebe, R. Suryanarayanan, G. Dhaleen, A. Revcolevschi, Phys. Rev. B **61**, 570 (2000).
12. J.P. Hannon, G.T. Trammell, M. Blume, D. Gibbs, Phys. Rev. Lett. **61**, 1245 (1988).
13. J.P. Hannon, G.T. Trammell, M. Blume, D. Gibbs, Phys. Rev. Lett. **62**, 2644 (1989).

14. For a review see *Correlated Electron Systems*, Science **288**, 461 (2000), and references therein.
15. P.G. Radaelli, D.E. Cox, M. Marezio, S-W. Cheong, Phys. Rev. B **55**, 3105 (1997).
16. S. Mori, C-H. Chen, S-W. Cheong, Nature **392**, 473 (1998).
17. T. Mutou, H. Kontani, Phys. Lett. **83**, 3685 (1999).
18. D. Khomskii, J. van den Brink, Phys. Rev. Lett. **85**, 3329 (2000).
19. T. Hotta, E. Dagotto, H. Koizumi, Y. Takada, Phys. Rev. Lett. **86** 2478 (2001).
20. A. Daoud-Aladine, J. Rodriguez-Carvajal, L. Pinsard-Gaudart, M.T. Fernandez-Diaz, A. Revcolevschi, Phys. Rev. Lett. **89**, 097205 (2002).
21. Y. Moritomo, Y. Tomioka, A. Asamitsu, and Y. Tokura and Y. Matsui, Phys. Rev. Lett. **51**, 3297 (1995).
22. B.J. Sternlieb, J.P. Hill, U.C. Wildgruber, G.M. Luke, B. Nachumi, Y. Moritomo, Y. Tokura, Phys. Rev. Lett., **76**, 2169 (1996).
23. Y. Murakami, H. Kawada, H. Kawata, M. Tanaka, T. Arima, Y. Moritomo, Y. Tokura, Phys. Rev. Lett. **80**, 1932 (1998).
24. J.B. Goodenough, Phys. Rev. **100**, 564 (1955).
25. E.O. Wohlan, W.C. Koehler, Phys Rev. **100**, 545 (1955).
26. J.P. Hill, D.F. McMorrow, Acta. Crystallogr. A **52**, 236 (1996).
27. M. Blume, J. Appl. Phys. **57**, 3615 (1985).
28. T. Mizokawa, A. Fujimori, Phys. Rev. B **56**, R493 (1997).
29. I.S. Elfimov, V.I. Anisimov, G.A. Sawatzky, Phys. Rev. Lett. **82**, 4264 (1999).
30. M. Benfatto, Y. Joly, C.R. Natoli, Phys. Rev. Lett. **83**, 636 (1999).
31. I.V. Solovyev, K. Terakura, Phys. Rev. Lett. **83**, 2825 (1999).
32. S. Larochelle, A. Mehta, N. Kaneko, P.K. Mang, A.F. Panchula, L. Zhou, J. Arthur, M. Greven, Phys. Rev. Lett. **87**, 095502 (2001).
33. P. Mahadevan, K. Terakura, D.D. Sarma, Phys. Rev. Lett. **87**, 066404 (2001).
34. P. Benedetti, J. van den Brink, E. Pavarini, A. Vigliante, P. Wochner, Phys. Rev. B **63**, 060408 (2001).
35. S.B. Wilkins, P.D. Spencer, P.D. Hatton, S.P. Collins, M.D. Roper, D. Prabhakaran, A.T. Boothroyd, Phys. Rev. Lett. **91**, 167205 (2003).
36. C.W.M. Castleton, M. Altarelli, Phys. Rev. B **62**, 1033 (2000).
37. S.S. Dhesi, A. Mirone, C. De Nadai, P. Ohresser, P. Bencok, N.B. Brookes, P. Reutler, A. Revcolevschi, A. Tagliaferri, O. Toulemonde, G. van der Laan, Phys. Rev. Lett. **92**, 056403 (2004).
38. Y. Murakami, J.P. Hill, D. Gibbs, M. Blume, I. Koyama, M. Tanaka, H. Kawata, T. Arima, Y. Tokura, K. Hirota, Y. Endoh, Phys. Rev. Lett. **81**, 582 (1998).

# Spin Switching in Mesoscopic Ring Magnets

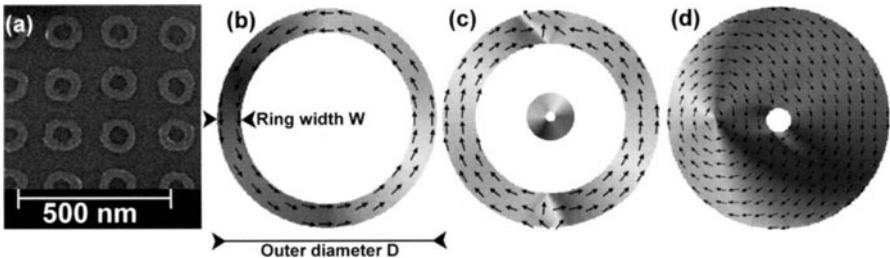
M. Kläui

Fachbereich Physik, Universität Konstanz  
Universitätsstr. 10, 78457 Konstanz, Germany

**Abstract.** In this article a comprehensive investigation of the spin switching in ferromagnetic NiFe and Co ring magnets is reported. Three types of switching (single, double, triple) are identified and the switching type is determined systematically as a function of the geometrical parameters in a phase diagram. The switching types and the switching field variation with varying ring diameter, ring width and film thickness are studied and explained in terms of the different physical processes occurring during the transitions (domain wall depinning, propagation, nucleation, etc.). By adjusting the ring geometry, the switching fields can be tailored over a wide range of values spanning more than an order of magnitude.

## 1 Introduction

The physics of surfaces, interfaces and thin films has become one of the main areas of research in the last few decades, due to the trend in science and technology towards miniaturisation of physical systems into the nanoscale. The miniaturisation trend demands as complete an understanding as possible of the physical properties of systems whose boundaries constitute an appreciable portion of the overall system [1,2]. From a technological point of view, some of the properties exclusive of mesoscopic systems (thin films, multilayers, fine particles, patterned elements, etc.) hardly need to be emphasized, so obvious and conspicuous have been their practical applications, and this is particularly true of magnetism (e.g. sensors, storage, etc. [3]). Understanding and controlling the magnetic properties of small ferromagnetic elements is a major challenge in the rapidly evolving field of nanoscale science, since these structures allow for the investigation of fundamental physical properties and also because they have important potential applications such as in magnetic random access memory (MRAM) cells [4]. One key issue is to understand and control the magnetic switching precisely. To achieve this one needs firstly to have well defined and reproducible remanent states and secondly, the switching process itself must be simple and reproducible. A geometry that fulfills these criteria is the ring geometry, which has recently become the focus of intense interest [5,6,7,8,9,10,11]. In addition to the quasi-static switching using externally applied fields, recently switching in rings by current induced domain wall propagation [12] and ultra-fast precessional switching by field pulses [13,14] was studied as well. The ring geometry is not just of interest for device applications, such as magnetic random access memory [4] but it is



**Fig. 1.** (a) SEM image of part of an array of NiFe rings (outer diameter  $D = 110$  nm, ring width  $W = 25$  nm, film thickness  $t = 10$  nm). (b-d) Micromagnetic simulations of the different magnetic states: vortex state (b), onion state (c), vortexcore state (d). The definition of the geometrical parameters outer diameter  $D$  and ring width  $W$  are shown in (b) and the arrows and the colour wheel in (c) give the magnetization direction

also particularly suited for probing fundamental magnetic properties such as domain wall trapping [15] and spin switching.

Using externally applied fields, so far three types of switching between different magnetic states present in rings have been identified [10,11]: (i) A double switching exhibited by most rings [5,16], where in addition to the flux-closure vortex state (Fig. 1 (b)), the high-moment ‘onion’ state (Fig. 1 (c)) is observed, which is characterized by two head-to-head domain walls and is attained reversibly from saturation. Thus in this case a ring switches first from the onion state to the vortex state and then to the opposite onion state [6]. (ii) A single switching typically observed in thin rings, which corresponds to switching from the onion to the reverse onion state [11] and (iii) a triple switching that occurs in very wide and thick rings, where the ring switches from the onion to the vortex state and then from the vortex state to the ‘vortexcore’ state [11], in which a complete vortex core is present in the ring (Fig. 1 (d)). Finally the vortex core is pushed out of the ring and the transition to the opposite onion state occurs. While there are a few isolated studies on the switching processes [5,16,11,10,17], the physical reasons for the dependence of the type of switching on the geometry have not been explained so far. What is more, the geometrical phase boundaries between the different types of switching behaviour can only be properly understood if the dependence of the transition fields on the geometry is known. Only a systematic study comparing all the switching fields for a single material can reveal the energetics governing the transitions. A study of the switching field range, which is achievable, is also of paramount importance for possible use in applications, since the fields needed to reverse e.g. a memory cell are limited by the possible geometries of the field generating striplines.

In this paper the recent work on spin switching in ferromagnetic rings is reported. In particular, a comprehensive systematic study of the magnetization switching as a function of the geometrical parameters is presented. This

includes the switching phase diagram for the type of switching as a function of the ring geometrical parameters and a detailed account of the switching fields for the different transitions. Arrays of polycrystalline Co and Permalloy (NiFe) rings were fabricated as detailed in [17] and a scanning electron microscopy image is shown in Fig. 1 (a). The type of switching and the switching fields were determined by magneto-optical Kerr effect with the magnetic field applied in the plane of the ring [17].

## 2 Switching Type Phase Diagram

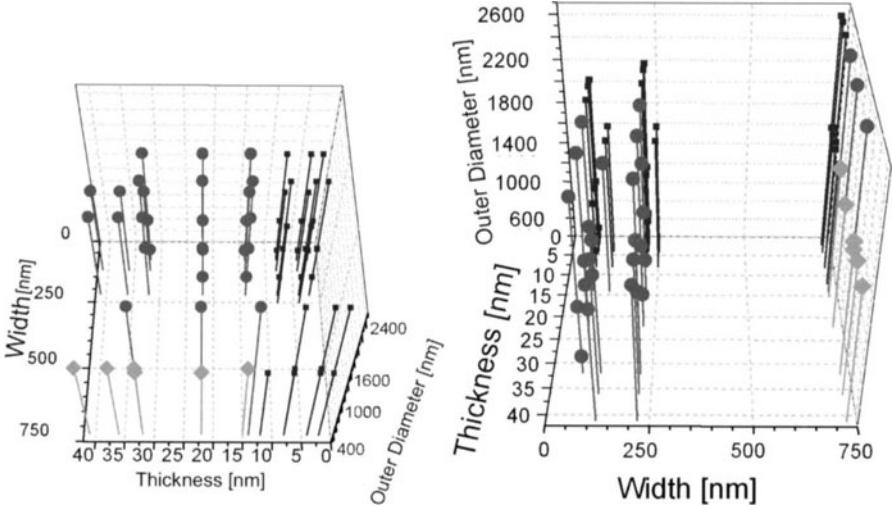
Qualitatively three different types of switching are observed as discussed above (single, double, triple). By analysing rings with more than 70 different combinations of the geometrical parameters it has been possible to obtain a switching phase diagram where, as a function of outer diameter, ring width and thickness, the rings are colour coded according to their switching behaviour. This four-dimensional information is plotted in a 3D plot with the type of switching colour coded in the symbols. The 3D plot is shown from different perspectives depending on the parameter of interest.

### 2.1 Thickness Variation

First the dependence on the thickness (phase diagram 3D plot in Fig. 2 (a)) is investigated. Most obvious is the sharp boundary between the single switching and the other two switching types between  $t = 8 \text{ nm}$  and  $t = 15 \text{ nm}$ . Thus the main geometrical factor that determines the switching behaviour of the ring magnets is the film thickness, which sets a sharp separation between these two switching regimes and consequently multiple steps in the  $M$ - $H$  loop are present in the thicker rings while thinner rings show single transitions from the onion to the reverse onion state. For very thin rings the stray field energy is reduced with respect to the exchange energy, making the vortex state unfavourable so that it might become unstable. Even if the vortex state in thin rings is stable at remanence, it will not be attained, since the reversal does not take place by the propagation of a head-to-head domain wall but by the energetically more favourable nucleation and expansion of a reverse domain somewhere in the ring [10,11]. In wider rings, the spins are less strongly aligned to the circumference and hence can twist more easily to form the core of a reverse domain and thus as reported in [17] the single switching extends to slightly larger thicknesses in wide rings than in narrow rings. Most importantly it should be noted that the thickness is the dominating parameter for the transition from single to multi-step switching.

### 2.2 Width Variation

In Fig. 2 (b) the switching field diagram oriented towards the width-axis is presented. It can be seen that for sufficiently wide and thick rings triple



**Fig. 2.** Phase diagram of the type of switching (single switching: blue squares; double switching: red circles; triple switching: green diamonds) for polycrystalline Co rings. Left: (a) The diagram is shown from a perspective where the thickness dependence can be easily discerned. There is a clear transition from the single switching to a multi-step switching with a sharp boundary between 8 nm and 15 nm. Right: (b) The diagram is shown from a perspective where the width and outer diameter dependence can be easily discerned. A transition from the double to the triple switching can be observed along the width and along the outer diameter axes

switching occurs, while for narrow and sufficiently thick rings double switching takes place. This is easy to understand, since for the vortexcore state to be energetically favourable, the ring has to be wide enough to accommodate a complete vortex core and also thick enough to make the flux closure state more favourable than a state involving a stray field. For rings that are wide, thick and have a large outer diameter, double switching occurs again, which will be explained in the next section.

### 2.3 Outer Diameter Variation

In Fig. 2 (b) the dependence of the type of switching on the outer diameter can be seen as well. 700 nm wide rings exhibit a triple switching for outer diameter  $D_{\text{ext}}=1.7 \mu\text{m}$  whereas for  $D_{\text{ext}}=2.7 \mu\text{m}$  a double switching occurs even though the ring is wide enough to accommodate a vortex core. To understand for which geometries the vortex to vortexcore transition occurs (which leads to a triple switching) and when the vortex to onion transition occurs (which leads to the double switching), we have to take into account the fact that the nucleation of the vortex core involves strong twisting of the spins at the inner edge to form the circular vortex structure [11]. This is facilitated by a strong curvature of the inner edge, since the spins already form a curved

structure and it is easier to twist them to form the complete vortex. For constant thickness and width, smaller rings have a higher curvature leading to more twisting of the spins along the circumference which leads to an easier nucleation of a vortex core.

## 2.4 Switching Type Phase Diagram for Multi-step Switching

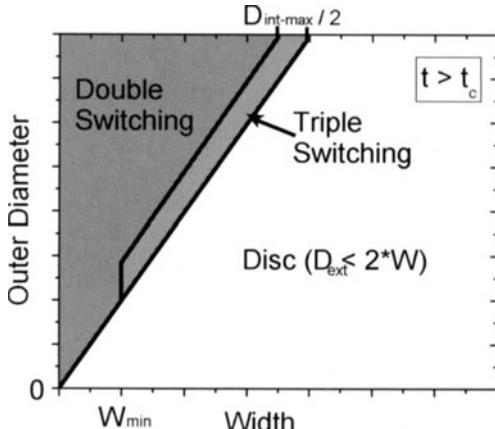
The explanations of the different switching processes are backed up by results from experiments and micromagnetic simulations and conceptually this then leads to the qualitative phase diagram (of rings with a thickness above the threshold where multi-step switching occurs) shown in Fig. 3: If the outer diameter  $D_{\text{ext}}$  is equal to  $2 \times W$  (width) then the element is a disc rather than a ring (black line); the relevant part of the diagram, is the part above the line, which corresponds to rings. For small widths  $W < W_{\min}$  the ring is too narrow to accommodate the vortex core and hence a double switching occurs ( $W_{\min}$  is to some extent thickness and material dependent). If the outer diameter is only slightly larger than  $2 \times W$  then this corresponds to a small central hole and hence a high curvature of the inner edge and so triple switching occurs. If the outer diameter is much larger than  $2 \times W$ , then the central hole is large; in this case, there is only little curvature of the inner edge and hence in this case a double switching process occurs. From this it can be seen that in addition to the width  $W$ , the inner diameter plays a crucial role for the double to triple switching transition with triple switching occurring for rings with sufficiently small inner diameters  $0 < D_{\text{int}} < D_{\text{int}}^{\max}$  (but only in sufficiently wide rings).

## 3 Switching Field Values

Having established qualitatively, which switching processes occur for which geometrical parameters, the next step is the investigation of the quantitative evolution of the different switching field values as a function of the geometrical parameters, which will in turn allow an understanding of the energetics that determine the phase diagram.

### 3.1 Outer Diameter Variation

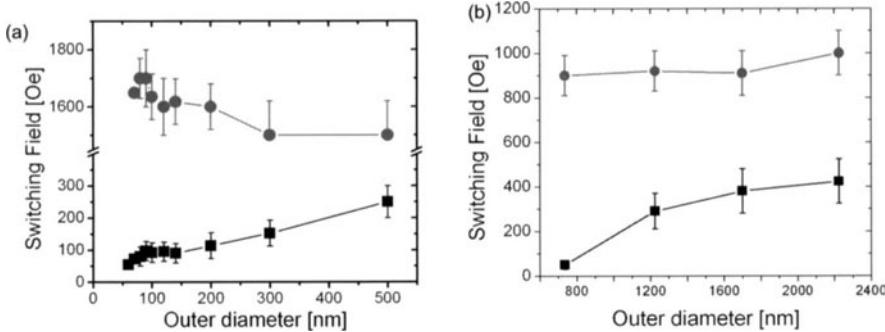
For the systematic investigation of the switching field values first the dependence of the switching fields on the lateral size is studied as ring width and film thickness are kept constant; this essentially is an investigation of the influence of the curvature on the switching and to investigate this to the smallest dimensions the smallest reproducible linewidth (25 nm) is used, which means that rings down to 90 nm outer diameter can be investigated (limited by the fabrication). This is also of importance for applications as the memory density is expected to soon reach values where each memory



**Fig. 3.** Qualitative phase diagram of the multi-step switching region (thickness above the critical thickness  $t_c$  where the transition from single to multi-step switching occurs, and lateral dimensions below the threshold where defect dominated multi-domain states occur) [20]. It can be seen that the triple switching only occurs in a limited region, where the ring is wide enough to accommodate a vortex core and where the inner diameter is small enough so that the curvature of the inner edge is large

cell has sub-100 nm dimensions. Conceptually one would expect that as the lateral size of the rings is reduced, the vortex state becomes more and more unfavourable due to the increasing curvature of the ring, which results in an increased exchange energy. Since it is the stray field-free vortex state that has been suggested for applications in data storage [4], it is of paramount importance to establish whether the double switching process, where the vortex state is attained, exists in the smallest rings, and how the switching fields develop when the lateral dimensions are reduced. To check this, a soft magnetic material (permalloy) is used for the rings. It is found that even for the smallest rings, double switching can be obtained if a sufficiently thick film, e.g. the 10 nm of permalloy used here [18].

Fig. 4 (a) shows the onion to vortex (black squares) and vortex to reverse onion (red circles) switching fields as a function of outer diameter for rings with a constant ring width (25 nm) and thickness (10 nm Permalloy). The high switching field of the vortex to reverse onion transition can be explained by the fact that the ring width is narrow (this dependence of the switching field on the width has also been observed in Ref. [17]). The vortex to reverse onion switching field is only weakly dependent on the outer diameter (the relative change is only about 10%), while the onion to vortex switching field increases significantly with increasing outer diameter (the switching field more than triples). This behaviour is due to a domain wall depinning and propagation process [5]: The effective field working to displace the wall is the field component that is parallel to the direction of the domain wall



**Fig. 4.** Switching field trend for the onion to vortex (black squares) and the vortex to reverse onion (red circles) transitions for rings of a constant width and thickness: (a)  $W = 25 \text{ nm}$  and  $t = 10 \text{ nm}$  polycrystalline Py; (b)  $W = 250 \text{ nm}$  and thickness  $t = 32 \text{ nm}$  polycrystalline Co. The error stems primarily from the uncertainty of extracting the average switching field due to the inherent switching field distribution of the arrays measured. The onion to vortex transition exhibits an increasing switching field with increasing outer diameter, whereas the vortex to reverse onion transition is largely unaffected

movement, hence the component parallel to the perimeter. If the domain wall positions are perfectly aligned with the field direction, this component is zero (field and domain wall positions are antiparallel). Due to local imperfections, the domain wall will be always displaced a few nanometer from the position along the field direction. In the case of a small outer diameter, due to the large curvature such a small displacement will lead to a reasonably large angle between the field and the domain wall position. Thus the field component parallel to the propagation direction (which is acting on the domain wall to move it) becomes large. In a ring with a large outer diameter, such a small displacement will mean that the angle and thus the effective field component are very small and thus it takes a higher (switching) field to move the wall. A similar dependence is found for the other rings studied. As an example in Fig. 4 (b) the dependence is shown for larger rings with a thick polycrystalline Co film and the same trend is observed as for the Py rings.

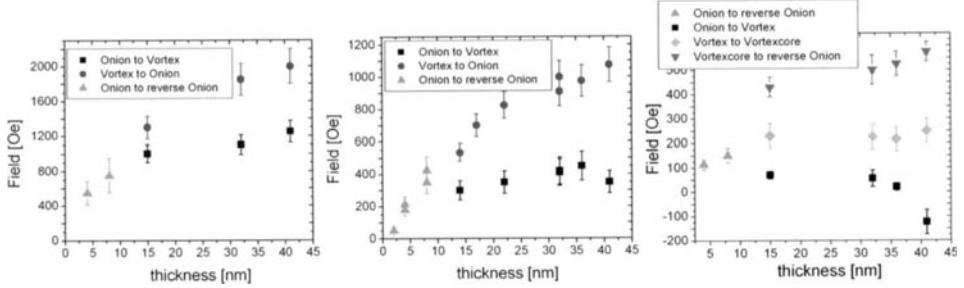
A useful consequence of the different dependences of the different switching fields on the lateral size is the fact that as the lateral size is reduced the switching fields from onion to vortex and vortex to reverse onion are increasingly separated. So even if a switching field distribution exists it is easy to reliably switch all the rings into the onion or vortex state as the two switching fields are well separated for small rings. For very large rings the switching field distributions might overlap, in which case it is not possible to switch all the rings into the desired state together. It remains to be noted that the single onion to reverse onion transition occurring in thin rings was found to

be only weakly dependent on the lateral size since it relies on the nucleation of a reverse domain.

### 3.2 Thickness Variation

As pointed out earlier a key factor influencing the type of switching is the film thickness. Here the switching field change with thickness is investigated, whilst the other geometrical parameters are kept constant. In Fig. 5 (a) the dependence of the switching fields on the thickness is presented for rings of 1.7  $\mu\text{m}$  outer diameter and 110 nm width. It can be seen that both the onion to vortex and the vortex to onion switching fields increase with increasing thickness. The increase for the onion to vortex transition (domain wall depinning and propagation process) can be explained by the fact that the domain walls are more strongly pinned in thick films. Due to the increased stray field of the onion state in thick film rings, the domain walls are more sharply defined (e.g. the transverse domain wall is narrower) and more prone to pinning at non-magnetic defects in the film and at edge defects to reduce their energy. Since the vortex state becomes energetically more favourable for thick films, the field needed to switch the ring into the reverse onion state increases. It is interesting to note that the slope of the change of switching field with thickness is higher for the vortex to reverse onion transition. This means that at a critical thickness the two switching fields approximately equalize and at that thickness the ring would switch directly from one onion to the reverse onion state and the intermediate vortex state would not be attained anymore. The onion to reverse onion switching fields decrease with decreasing film thickness as the rings become "super soft" and the spins are less strongly aligned with the perimeter.

In Fig. 5 (b) a similar graph is presented for rings with 250 nm width, where more data points are available. The same behaviour as discussed above is found apart from the fact the onion to vortex transition switching field starts to decrease at large thicknesses. This is even more obvious in Fig. 5 (c) where for 700 nm wide rings it can be seen that the onion to vortex switching field decreases with increasing thickness (negative switching fields correspond to switching before zero Oe when lowering the field). In the onion state in wide rings the domain walls interact more strongly with each other since their size increases with increasing width. Because the pinning is much less than in narrow rings, the domain walls may annihilate one another and create the energetically favourable vortex state in thick rings when the applied field is lowered before zero Oe is reached. And since the vortex state becomes more and more favourable for increasing thickness, the transition field decreases. The vortex to vortexcore transition is not significantly affected by the thickness, since both states are stray field-free and hence the energy of both states scales equally with thickness. The vortexcore to reverse onion transition field increases with increasing thickness, since the stray field-free vortexcore state



**Fig. 5.** Switching fields of polycrystalline Co rings with 1.7  $\mu\text{m}$  outer diameter as a function of thickness for different widths: Left: (a)  $W = 115 \text{ nm}$ ; thin rings show a single switching and thick rings a double switching with all switching fields increasing with increasing thickness. Middle: (b)  $W = 250 \text{ nm}$ ; Thin rings show a single switching and thick rings a double switching with the switching fields generally increasing with increasing thickness. Right: (c)  $W = 700 \text{ nm}$ ; Thin rings show a single switching and thick rings a triple switching. The onion to vortex switching field decreases, whereas the vortexcore to reverse onion switching field increases with increasing thickness

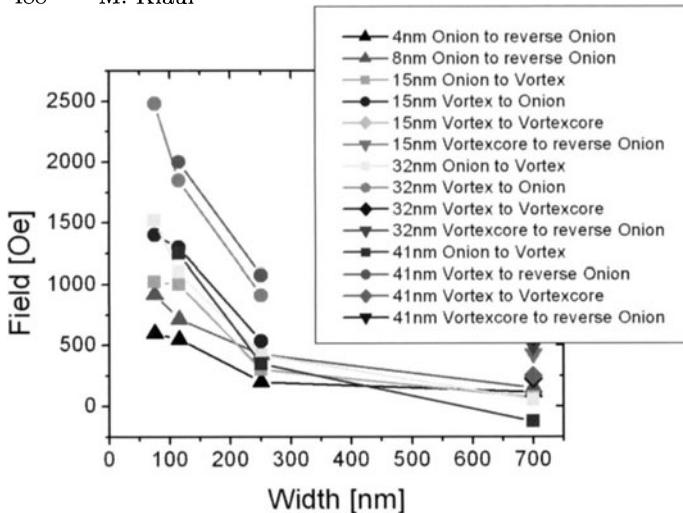
is more favourable for thick films as seen in the highest field transition for all switching types.

### 3.3 Width Variation

For epitaxial fcc Co rings a detailed study of the width dependence of the vortex to onion switching field was carried out (see in Ref. [16]). Polycrystalline Co rings were considered above and as seen in Fig. 6 the general trend for all transitions and for all thicknesses is an increase in the switching fields with decreasing ring width, which was observed for the vortex to onion transition also in [16] and predicted from micromagnetic simulations [19]. Furthermore the dependence of the switching field on the ring width becomes stronger for narrower rings and also for larger thicknesses.

For the vortex to onion transition, this behaviour is qualitatively easy to understand as the vortex state is more stable in narrower rings. To understand this dependence better, the dynamics of the vortex to onion transition have to be studied. For narrow rings the simulations predict the transition to start with a nucleation of a reverse domain in the ring (as discussed in detail in Ref. [16]). This nucleation requires a large twisting of the spins at the nucleation site, which is harder to achieve in narrow rings where the spins are more strongly aligned. Comparing the experimental data with the numerical simulations shows good agreement for the switching fields of wide rings. For narrower rings the computations predict switching fields that are too large as detailed in [16,19].

The onion to reverse onion transition in thin rings seems to be the least dependent on the width but still shows a considerable increase of the switch-



**Fig. 6.** Switching fields of polycrystalline Co rings of  $1.7\text{ }\mu\text{m}$  outer diameter and different thicknesses as a function of width. There is a strong increase in all the switching fields with decreasing ring width. For reasons of clarity the error bars have been omitted but they are similar to those presented in previous graphs with switching field measurements

ing field with decreasing width. Since this transition is similar to the vortex to reverse onion transition (nucleation and spread of a reverse domain process), the explanation of this behaviour is similar with the spins in narrower rings resisting being twisted more because of the stronger alignment along the perimeter of the ring (local shape anisotropy).

The onion to vortex transition takes place at higher fields for narrower rings, since the pinning of the domain walls is stronger the narrower the rings. The narrowing of the ring width leads to an increase in the local (shape) anisotropy, which affects the width and shape of the domain wall and generally acts in such a way as to increase the switching field (e.g., the reduction in the domain wall length leads to a reduction in the wall pressure exerted by the applied magnetic field) [17]. In a narrow ring the increasing energy from the domain walls and the change in shape as they are squeezed into rings with a smaller width, contribute to an increase in the switching field. Finally, the increase in the switching field with decreasing ring width is also due in part to the effect of edge roughness, which acts as pinning sites for the domain wall and which becomes more pronounced for narrower rings. This mechanism is similar to that responsible for the increase in switching field with increasing film thickness in polycrystalline films.

## 4 Conclusions

In this article the results of an investigation into the spin switching in ferromagnetic rings are reported. The type of switching (single, double, triple) and the switching fields are explored for varying geometrical parameters in polycrystalline Co and Py rings. Thin rings exhibit a single onion to reverse onion transition, whereas thick rings show multi-step switching. For sufficiently wide and thick rings with a sufficiently small central hole triple switching is observed.

The onion to reverse onion switching field increases with increasing thickness and decreasing width, which both lead to a stronger alignment of the spins with the perimeter, hindering the nucleation of a reverse domain and it is not significantly affected by the outer diameter. The switching field values ranges from 50 Oe in the widest and thinnest case to just above 1 kOe for the narrowest and thickest rings (where this single jump switching occurs).

The same trend is observed for the vortex to reverse onion and the vortex-core to reverse onion switching fields. The vortex to reverse onion switching field ranges from less than 400 Oe to about 2.5 kOe for decreasing width and increasing thickness.

In narrow rings exhibiting a double switching the onion to vortex switching field increases with decreasing width and at first increases with thickness up to a maximum but can decrease again as the thickness is increased further (depending on the exact width), which is explained by the dependence of the domain wall depinning and propagation (responsible for this transition) on the width and thickness. For rings that exhibit a triple switching this transition can occur when lowering the field before reaching zero Oe (at fields below  $-125$  Oe) and in this case the switching field actually decreases with increasing thickness, since the vortex state becomes more favourable. Concerning the dependence on the outer diameter, in general the switching field increases with increasing outer diameter and this switching field can range from less than  $-125$  Oe to more than 1.5 kOe.

In conclusion it is shown that the various switching processes and field values depend strongly on the different geometrical parameters and this dependence can be explained by the physical processes occurring. By varying the geometry, the switching fields can be tailored to a wide range of values spanning more than an order of magnitude.

## Acknowledgements

This work has been supported by the EU network MASSDOTS, the CMI program ‘Magneto-electronic devices’, the Gottlieb Daimler and Carl Benz foundation, the ‘Deutsche Forschungsgemeinschaft’ (SFB 513) and the ‘Deutscher Akademischer Austausch Dienst’. I would like to thank U. Rüdiger, J. A. C. Bland, C. A. F. Vaz, L. J. Heyderman and L. Lopez-Díaz for very fruitful collaborations.

## References

1. W. Kohn, Rev. Mod. Phys. **71**, S59 (1999)
2. F. J. Himpsel et al., Adv. Phys. **47**, 511 (1998)
3. G. Prinz, J. Magn. Magn. Mater. **200**, 57 (1999)
4. J.-G. Zhu et al., J. Appl. Phys. **87**, 6668 (2000)
5. J. Rothman et al., Phys. Rev. Lett. **86**, 1098 (2001).
6. M. Kläui et al., Appl. Phys. Lett. **78**, 3268 (2001).
7. M. Kläui et al., Appl. Phys. Lett. **81**, 108 (2002).
8. S. P. Li et al., Phys. Rev. Lett. **86**, 1102 (2001).
9. S. Kasai et al., J. Magn. Magn. Mater. **239**, 228 (2002).
10. M. Kläui et al., J. Phys., Condens. Matter. **15**, 985 (2003).
11. M. Kläui et al., Phys. Rev. B **68**, 134426 (2003).
12. M. Kläui et al., Appl. Phys. Lett. **83**, 105 (2003).
13. L. Lopez-Diaz et al., J. Appl. Phys. **89**, 7579 (2001).
14. L. Lopez-Diaz et al., Physica B **306**, 211 (2001).
15. M. Kläui et al., Phys. Rev. Lett. **90**, 97202 (2003).
16. M. Kläui et al., J. Magn. Magn. Mat. **240**, 7 (2002).
17. Y. G. Yoo et al., Appl. Phys. Lett. **82**, 2470 (2003).
18. L. J. Heyderman et al., J. Appl. Phys. **93**, 7349 (2003).
19. L. Lopez-Diaz et al., IEEE Trans. Magn. **36**, 3155 (2000).
20. M. Kläui et al., (unpublished).

# Modified Magnetic Properties of Paramagnetic (Zn,Mn)S at Reduced Dimensions

P. J. Klar<sup>1</sup>, L. Chen<sup>1</sup>, W. Heimbrodt<sup>1</sup>, F.J. Brieler<sup>2</sup>, M. Fröba<sup>2</sup>, T. Kurz<sup>3</sup>, H.-A. Krug von Nidda<sup>3</sup>, and A. Loidl<sup>3</sup>

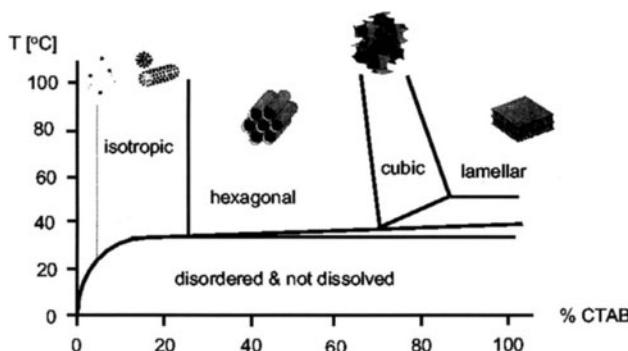
<sup>1</sup> Department of Physics and Materials Science Center, Philipps-University Renthof 5, 35032 Marburg, Germany

<sup>2</sup> Institute of Inorganic and Analytical Chemistry, Justus-Liebig University Heinrich-Buff-Ring 58, 35392 Giessen, Germany

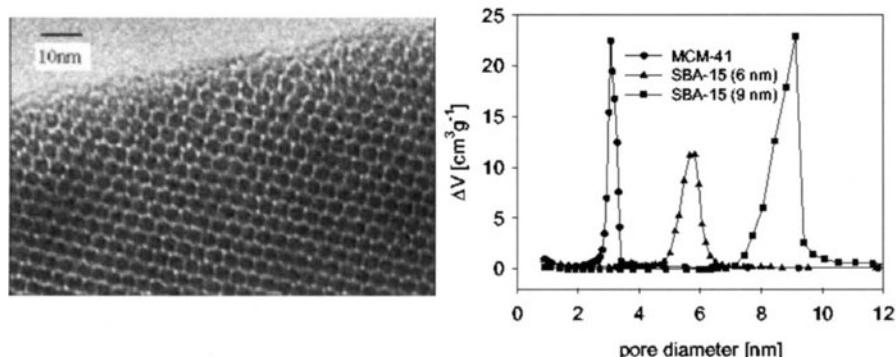
<sup>3</sup> Experimentalphysik V, EKM, Department of Physics, University of Augsburg Universitätsstraße 2, 86135 Augsburg, Germany

**Abstract.** We present a novel way of synthesizing highly ordered arrays of magnetic nanostructures with lateral dimensions of 2 to 10 nm by incorporation into mesoporous SiO<sub>2</sub> matrices. Such nanostructure arrays are suitable for studying magnetic phenomena at reduced dimensions, which is also important in the context of device miniaturization. Exemplarily we study (Zn,Mn)S nanowires. Changes of the macroscopic observables (e.g. Curie-Weiss parameter  $\Theta$  and line width  $\Delta H$  of the electron paramagnetic resonance) of the paramagnetic phase due to reduced dimensions are observed. The microscopic coupling between the Mn-ions (e.g. the exchange constants  $J_{nn}$  and  $J_{nnn}$ ) is not altered to a first approximation. The macroscopic modifications arise mainly due to geometrical restrictions, i.e. the number of neighbors in the cation shells around a Mn-ion in the surface region are considerably reduced compared to a Mn-ion in the bulk. Similar effects are also anticipated for antiferromagnetic and ferromagnetic wires.

Our bottom-up approach of fabricating high-density ordered arrays of regular magnetic nanostructures by incorporating them into the regular pores of SiO<sub>2</sub> matrices is very promising for studying magnetic properties of magnetic semiconductors at reduced dimensions in the regime between 2 and 30 nm. The high density of nanostructures does not allow probing them individually, but the amount of similar magnetic nanostructures is sufficient to employ almost any magnetic measurement technique. Furthermore, the well-defined pore diameter provides a sharp upper limit for the lateral extension of the nanostructures. Moreover, in contrast to the other methods for synthesizing quantum wires, the lateral dimensions become better defined with decreasing diameter of the structures. Here, we will discuss exemplarily, the effects of reduced dimensions on the magnetic properties of paramagnetic (Zn,Mn)S nanostructures. However, this approach is extendable to any magnetic material, which can be synthesized inside the SiO<sub>2</sub> pore systems, including ferromagnets and antiferromagnets.



**Fig. 1.** LC phases of the CTAB surfactant in  $\text{H}_2\text{O}$



**Fig. 2.** Left: High-resolution TEM image of a grain of a MCM-48  $\text{SiO}_2$  matrix. Right: Pore-size distributions of various mesoporous  $\text{SiO}_2$  matrices

## 1 Template-Directed Growth of Ordered Arrays of (II,Mn)VI Nanostructures

Mesoporous materials, suitable for fabricating such nanostructured host/guest compounds, were first introduced in 1992 [1]. To obtain the desired porous host structure with a high specific surface and a narrow, regular pore-size distribution it is necessary to use structure-directing agents in the synthesis. Commonly, amphiphilic molecules, e.g. low or high molecular weight surfactants, which are able to form lyotropic liquid crystals (LC) in the course of synthesis, serve as the respective structure-directors. Depending on the size and shape of these supramolecular templates mesoporous materials with well-defined pore diameters and structures (i.e. ordered arrays of nanotubes) can be obtained. Figure 1 depicts the LC phases obtained using the surfactant cetyltrimethylammoniumbromide (CTAB) as a function of surfactant concentration in  $\text{H}_2\text{O}$  and synthesis temperature. With different amphiphilic structure directing agents ordered mesoporous materials containing pores with diameters between 2 and 30 nm or more can be synthesized [2].

Removal of the organic template of the three-dimensional organic/SiO<sub>2</sub> superstructures (i.e. MCM-41S or SBA-15) in a calcination process leaves a highly ordered mesoporous SiO<sub>2</sub> matrix with regular wire-like pores. A high-resolution transmission-electron microscope (TEM) image of a grain of a MCM-48 SiO<sub>2</sub> matrix is shown in Fig. 2. The pore diameters as well as the distances between neighboring pores are only a few nanometers. Fig. 2 also shows pore-size distributions of three different hexagonal mesoporous silica derived from physisorption measurements. In particular, for the MCM-41 type silica the pore diameter is well defined. With increasing chain length the surfactant molecules show a tendency to bend and tangle up leading to a certain degree of disorder and the broadening of the pore distribution. The actual incorporation of the (II,Mn)VI and a thorough structural and electronic characterization of the semiconductor nanostructures is described in detail in Refs [3] and [4].

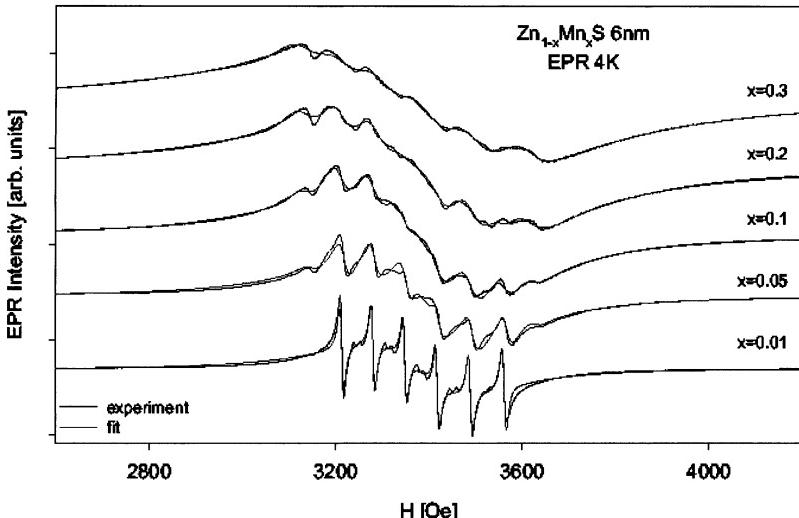
The hexagonal order of mesoporous MCM-41 and SBA-15 provides the possibility of growing dilute magnetic semiconductor quantum-wires inside the channels [5]. The incorporation of semiconductors into the pores is promising, because the SiO<sub>2</sub> wall structure with its large band gap serves as a barrier between single wires. Several conventional II-VI semiconductor compounds have been incorporated [6,7,8]. We have focussed on the incorporation of the magnetic semiconductors (Cd,Mn)S [5,9,3], (Cd,Mn)Se [9], (Zn,Mn)S [10,4].

## 2 Magnetic Properties

### 2.1 EPR Spectra

As in bulk (II,Mn)VI materials, the integral electron paramagnetic resonance (EPR) signal corresponding to the Mn<sup>2+</sup> absorption has a Lorentzian line shape in the paramagnetic regime [11]. Overall intensity  $I_{\text{tot}}$  as well as line width  $\Delta H$  of the Lorentzian are very sensitive to the spin-spin correlations between the Mn<sup>2+</sup> ions. Not only changes of the magnetic order with temperature (e.g. paramagnetic–spin-glass or paramagnetic–antiferromagnetic phase transitions) are reflected by their temperature dependence, but also changes of the magnetic interactions within the paramagnetic phase itself due to reduced lateral dimensions.

Figure 3 depicts normalized cw-EPR spectra of five 6 nm Zn<sub>1-x</sub>Mn<sub>x</sub>S wire samples with  $x = 0.01, 0.05, 0.1, 0.2$ , and  $0.3$  measured at 4 K with 9.5 GHz. These are characteristic for exchange coupled Mn<sup>2+</sup> ions in II-VI semiconductors [12]. The spectra basically consist of a sextet of sharp lines. This sextet centered at  $g = 1.999$  is associated with the allowed ( $\Delta m_S = \pm 1, \Delta m_I = 0$ ) magnetic dipolar transitions between the hyperfine-split Zeeman levels of the  $^6S_{5/2}$  ground state of the Mn<sup>2+</sup>  $3d^5$  electrons. The hyperfine structure arises from the interaction between the  $S = 5/2$  spin of the unpaired  $3d^5$  electrons with the  $I = 5/2$  spin of the  $^{55}\text{Mn}$  nucleus. The hyperfine splitting characteristic for Mn<sup>2+</sup> in ZnS amounts to about  $\delta B_{\text{HFS}} = 7.0$  mT between neighbor-



**Fig. 3.** EPR spectra of 6 nm  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wires with  $x$  of 0.01 to 0.3 at  $T = 4\text{ K}$

ing allowed transitions in zincblende as well as in wurtzite structure [13], in agreement with the splitting of the 1% sample. With increasing Mn-content the dipolar interaction and exchange coupling merge the hyperfine structure into one broad resonance line as for  $x = 0.3$ . But this broad line can already be identified for  $x = 0.01$ . The spectrum for  $x = 0.01$  is satisfactorily described by the sum of the broad line and the hyperfine structure of six lines. Moreover, at low  $x$  each hyperfine line exhibits a pair of satellites at lower external magnetic field associated with the forbidden ( $\Delta m_S = \pm 1$ ,  $\Delta m_I = \pm 1$ ) hyperfine transitions. The forbidden hyperfine transitions in the spectra with the lowest Mn-content ( $x = 0.01$ ) are typical for  $\text{Mn}^{2+}$  ions in the tetrahedral environment of a Zn-site in a zincblende crystal. For wurtzite crystals, these lines are much more prominent and merge with the allowed hyperfine transitions, as it is observed for samples with  $x \geq 0.05$ . Thus, the crystal structure of the incorporated  $(\text{Zn}, \text{Mn})\text{S}$  semiconductors is zincblende only for  $x = 0.01$  and wurtzite for higher  $x$ . Furthermore, an additional hyperfine structure of six lines with a larger splitting of about  $\delta B_{\text{HFS}} = 9.5\text{ mT}$  evolves with increasing  $x$  and persists even at 30% Mn-content. This hyperfine structure may result from isolated  $\text{Mn}^{2+}$  ions on the surface or interstitial places of the  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  nanowires [13]. For this particular set of samples we showed that the majority of the  $\text{Mn}^{2+}$  ions is well incorporated into the  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  nanowires and only a small amount remains weakly bound at their surface [4]. This amount of aggregated Mn-ions at the surface of the nanostructures corresponds to less than 4%, 2% and 1% of the total Mn-content for the 3, 6 and 9 nm wires, respectively, in agreement with the surface to volume ratios.

## 2.2 Analysis of the Curie-Weiss Parameter

The Curie-Weiss parameter  $\Theta$  of the paramagnetic phase is a measure for the type and strength of the interaction between the Mn-ions. In the mean field approximation,  $\Theta$  of a  $\text{II}_{1-x}\text{Mn}_x\text{VI}$  dilute magnetic semiconductor can be written as [14]

$$\Theta(x) = \frac{2}{3}S(S+1)x \sum_p \frac{J_p z_p}{k_B}, \quad (1)$$

where  $J_p$  is the exchange integral between  $p$ th neighbors and  $z_p$  is the number of cations in the  $p$ th coordination sphere. The Curie-Weiss parameter can be obtained experimentally either from the temperature dependence of the inverse EPR intensity  $I_{\text{tot}}^{-1}$  or the inverse susceptibility  $\chi^{-1}$ . At high temperatures, it holds for (II,Mn)VI:

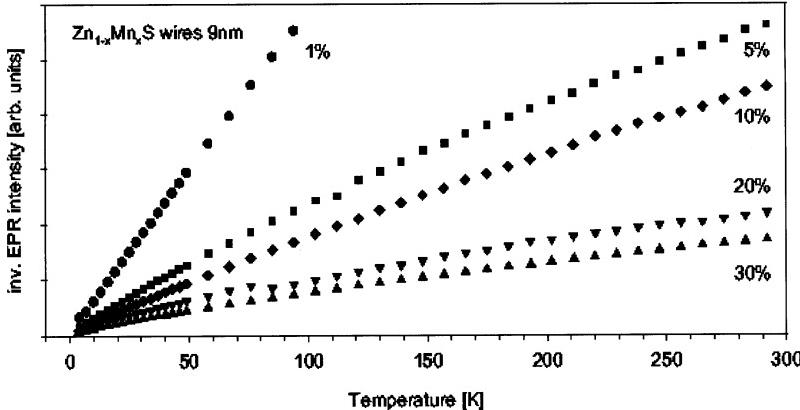
$$I_{\text{tot}}^{-1}(T) \propto \chi^{-1}(T) \propto (T + \Theta). \quad (2)$$

At low temperatures, typically below 200 K, deviations from the Curie-Weiss behavior occur when antiferromagnetic clustering between Mn-ions starts to play a role [15]. This is expected as the nearest-neighbor coupling  $J_1/k_B = J_{\text{nn}}/k_B \approx -10$  K. Further complications arise for  $x \geq 0.2$  because of the transition from the paramagnetic phase to the spin-glass phase at low temperatures [16,17]. It is found in the case of bulk (II,Mn)VI, that the experimentally determined Curie-Weiss parameter  $\Theta$  is usually well described, when the first (nn) and the second (nnn) shell of the Mn-ions are taken into account in Eq. 1 and higher shells are neglected [17], i.e.

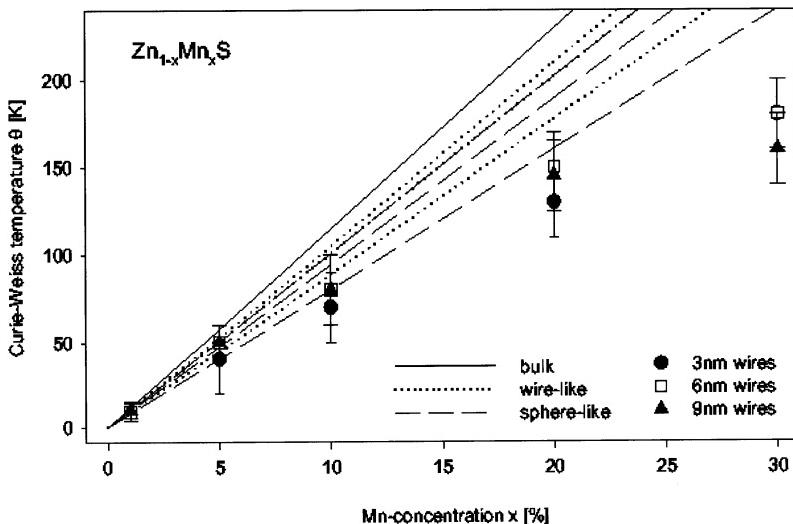
$$\Theta(x) = \frac{2}{3}S(S+1)x \left[ \frac{J_{\text{nn}}z_{\text{nn}}^{\text{b}}}{k_B} + \frac{J_{\text{nnn}}z_{\text{nnn}}^{\text{b}}}{k_B} \right]. \quad (3)$$

For wurtzite and zincblende crystals  $z_{\text{nn}}^{\text{b}}$  and  $z_{\text{nnn}}^{\text{b}}$  are 12 and 6, respectively.

Exemplarily, Fig. 4 shows plots of the inverse EPR intensity  $I_{\text{tot}}^{-1}$  of the 9 nm  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wire samples with  $x$  ranging from 1% to 30% as a function of temperature. As in the case of bulk (II,Mn)VI, the slope of the  $I_{\text{tot}}^{-1}(T)$  curve decreases with increasing Mn-content  $x$  and, at low temperatures, deviates from the Curie-Weiss behavior giving the curves a somewhat negative curvature. Fitting the  $I_{\text{tot}}^{-1}(T)$  plots in the high temperature regime (250 to 300 K) using the dependence in Eq. 2 yields the Curie-Weiss parameter  $\Theta$  as a function of  $x$ . Similar results were also obtained for the series of 3 and 6 nm  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wires. Plots of  $\Theta$  versus  $x$  obtained for  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wires of different diameters are summarized in Fig. 5. As expected,  $\Theta$  increases with increasing  $x$  in each series. The  $\Theta$ -values obtained for the nanostructures are considerably lower than those found in corresponding bulk (II,Mn)VI samples. The bulk dependence is represented by the solid lines in Fig. 5 and is calculated using Eq. 3. The exchange parameters from the literature are  $J_{\text{nn}} = -16.1$  K and  $J_{\text{nnn}} = -0.6$  K for (Zn,Mn)S [18]. Moreover, it appears



**Fig. 4.** Inverse EPR intensity  $I_{\text{tot}}^{-1}$  vs temperature  $T$  of the 9 nm  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wires



**Fig. 5.** Plots of the absolute value of the Curie-Weiss temperature  $\Theta$  vs Mn-content  $x$  for the  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wires. The solid line is calculated using Eq. 3, dashed and dotted lines are calculated using Eq. 5 for spherical and wire-like nanoparticles

that the  $\Theta$ -values show a tendency to decrease with decreasing wire diameter  $d$  at constant  $x$ . These effects are due to the reduced lateral dimensions of the nanostructures. They occur because Mn-ions in the outer layer of the (II,Mn)VI nanostructures incorporated inside the mesoporous  $\text{SiO}_2$  matrices have reduced numbers of nearest neighbors  $z_{\text{nn}}^{\text{s}} \approx z_{\text{nn}}^{\text{b}}/2$  and next-nearest neighbors  $z_{\text{nnn}}^{\text{s}} \approx z_{\text{nnn}}^{\text{b}}/2$  compared to bulk material. This becomes significant as the surface to volume ratio is strongly increased in the nanostructures. The effect can be estimated by dividing the volume  $V$  of the nanostructure into a volume  $V^{\text{s}}$  close to the surface (where the exchange effects differ from bulk)

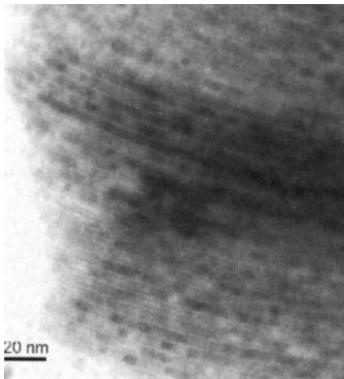
and a remaining bulk-like volume  $V^b = V - V^s$ . The choice of the two volumes will differ for nearest neighbors and next-nearest neighbors because the length scales involved, i.e. the nearest-neighbor distance  $d_{nn}$  and the next-nearest neighbor distance  $d_{nnn}$ , are different. In the following the wurtzite structure is approximated by a zincblende structure for simplicity. This is a good approximation here as only nearest and next-nearest neighbors are considered. It holds  $d_{nn} = \frac{1}{\sqrt{2}}a$  and  $d_{nnn} = a$  where  $a \approx 0.55$  nm is the lattice constant of zincblende ZnS. Two limiting cases for the shape of the nanostructure are considered: (i) an ideal wire structure of infinite length and diameter  $d$  and (ii) a spherical nanoparticle with diameter  $d$  where  $d$  is the pore diameter of the SiO<sub>2</sub> host matrix. One obtains the following definitions:

$$V_{nn}^s = V \left( 1 - \frac{(d - 2d_{nn})^\delta}{d^\delta} \right) \quad V_{nnn}^s = V \left( 1 - \frac{(d - 2d_{nnn})^\delta}{d^\delta} \right), \quad (4)$$

where  $\delta = 2$  for wires and  $\delta = 3$  for spheres. In both cases it holds  $V_{nn}^b = V - V_{nn}^s$  and  $V_{nnn}^b = V - V_{nnn}^s$ . Rewriting Eq. 3 including surface effects yields:

$$\Theta(x, d) = \frac{2S(S+1)x}{3Vk_B} [J_{nn}(V_{nn}^b z_{nn}^b + V_{nn}^s z_{nn}^s) + J_{nnn}(V_{nnn}^b z_{nnn}^b + V_{nnn}^s z_{nnn}^s)]. \quad (5)$$

It is assumed that the exchange parameters are the same in both volumes,  $V^s$  and  $V^b$ . Using Eq. 5 and the parameters given above, Curie-Weiss parameters have been calculated for the (Zn,Mn)S nanostructures assuming a wire-like as well as a spherical shape. The calculations for the Zn<sub>1-x</sub>Mn<sub>x</sub>S nanostructures were carried out for  $d$ -values of 3, 6 and 9 nm. The results are also plotted in Fig. 5. The dotted and dashed lines represent the results for a wire-like and a spherical shape, respectively. There are three calculated  $\Theta$ -curves for the wire-shaped and three calculated  $\Theta$ -curves for sphere-shaped nanoparticles. In both series the curves are assigned as follows to the corresponding  $d$ -values. The steepest curve corresponds to  $d = 9$  nm, the intermediate curve to  $d = 6$  nm and the lowest curve to  $d = 3$  nm. Comparing experimental data and the theoretical curves indicates that the latter describe the correct trends. The theoretically derived reduction of the Curie-Weiss parameter  $\Theta$  due to surface effects is, as expected, larger for spherical nanoparticles compared to wire-shaped nanoparticles. The calculated reductions are of the right magnitude, but still smaller than those found in the experiment. Assuming spherical particles yields a better agreement. This is in concordance with the TEM analysis of the (II,Mn)VI nanostructures. Figure 6 depicts a TEM image of a (Zn,Mn)S nanostructure incorporated into mesoporous SiO<sub>2</sub> with 6 nm pores. The TEM image is a cross sectional view. The narrow 'white' lines represent the SiO<sub>2</sub> walls and the dark regions are the (Zn,Mn)S nanoparticles. The aligned pore channels are filled with the (Zn,Mn)S compound. It can be seen that single 'sphere-like' nanoparticles are present, but also that these nanoparticles tend to agglomerate to denser 'wire-like' structures.



**Fig. 6.** TEM image of  $(\text{Zn},\text{Mn})\text{S}$  nanostructures incorporated into a SBA-15 mesoporous  $\text{SiO}_2$  matrix with a pore diameter  $d$  of 6 nm

### 2.3 Analysis of the EPR Line Width

The EPR line width of bulk  $(\text{Cd},\text{Mn})$ - and  $(\text{Zn},\text{Mn})$ -chalcogenides has been widely studied [11,19,20,21]. In such paramagnetic systems, its behavior depends intimately on the physics of the Mn spin-spin interactions. As a further complication in bulk  $(\text{II},\text{Mn})\text{VI}$  semiconductors with  $x \geq 0.2$  an additional broadening or even divergence of the EPR line width occurs in the vicinity of the paramagnetic spin-glass transition where the spin-spin correlation length diverges. None of the  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  and  $\text{Cd}_{1-x}\text{Mn}_x\text{S}$  wires with  $x \leq 0.3$  shows signs of the paramagnetic–spin-glass transition. The corresponding EPR line widths in Fig. 7 remain finite even at the lowest temperature. The critical Mn-concentration  $x$ , above which the phase transition occurs, corresponds to the percolation threshold for the Mn-ions on the cation sublattice. This percolation threshold is increased with the reduction of the dimensionality and, thus, the magnetic phase transition is suppressed in the wires. Therefore, the line-width behavior in the wires is solely determined by the spin-spin interactions within the Mn-subsystem.

Figure 7 depicts plots of the EPR line width  $\Delta H$  versus temperature for various  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wire samples of different  $x$  and  $d$ . Temperature trends of  $\Delta H$  for these samples can be explained qualitatively in the same fashion as for bulk  $(\text{II},\text{Mn})\text{VI}$ . At low temperatures, when nearest-neighbor Mn-ions have dimerised to antiferromagnetic pairs, dipolar broadening dominates for the remaining unpaired spins. With increasing temperature, an exchange narrowing effect is observed due to the nearest neighbor exchange at temperatures, when the thermal energy is sufficient to break up antiferromagnetically coupled nearest neighbor pairs. The two series of  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wires with diameters of  $d = 6$  and 9 nm are very similar. A strong broadening of the line width with  $x$  is observed at low temperatures. With increasing temperature exchange narrowing occurs. As expected, the narrowing effect becomes more pronounced with increasing  $x$ . The situation for the 3 nm  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wire samples is different at low temperature, where virtually the same line width is

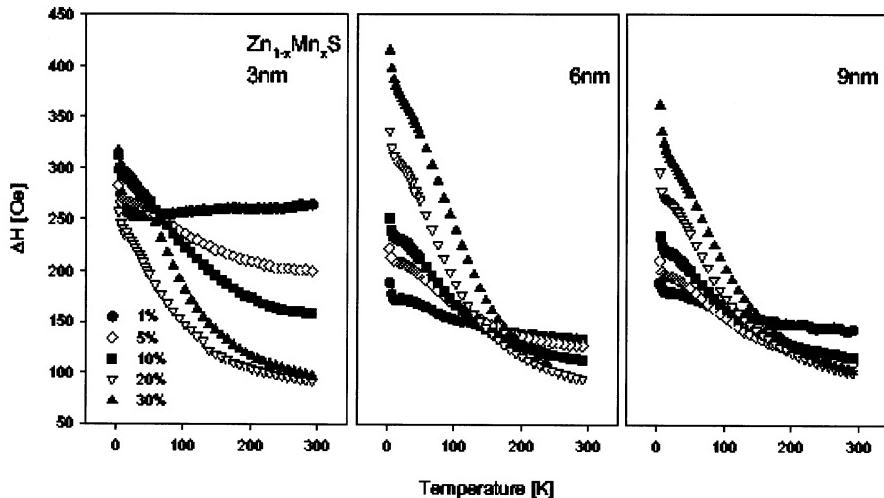


Fig. 7. Plots of the EPR line width  $\Delta H$  versus temperature for  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wire samples of different  $x \leq 0.3$  and diameters of 3 (left), 6 (middle), and 9 nm (right)

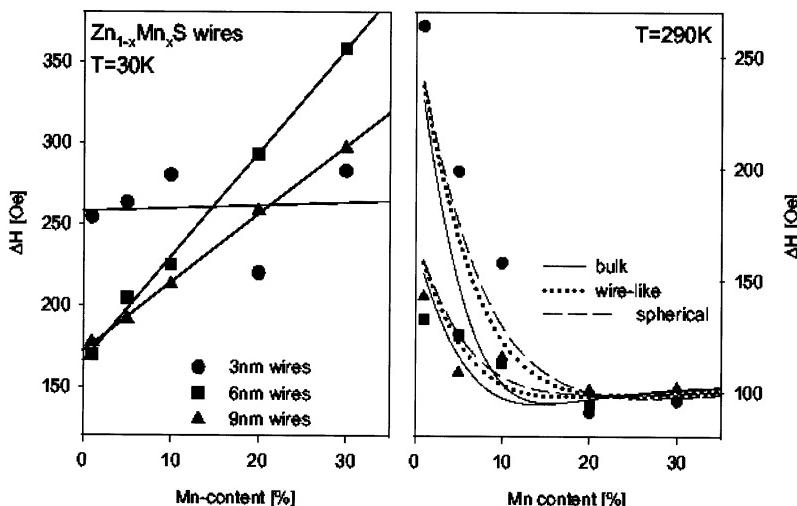


Fig. 8. Left: Plots of the EPR line width  $\Delta H$  as a function of  $x$  at  $T = 30$  K for  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wire samples of different diameters. The solid lines are linear fits. Right: Plots of the EPR line width  $\Delta H$  as a function of  $x$  at  $T = 290$  K for  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  wire samples of different diameters. The lines are model calculations

observed for all  $x$ . However, at high temperatures, strong exchange narrowing effects are also observed.

In the following, the concentration dependence of the line width at low temperatures ( $T = 30$  K) and at high temperatures ( $T = 290$  K) will be analyzed in more detail. Figure 8 shows plots of the line width at these tem-

peratures for the  $Zn_{1-x}Mn_xS$  wire series of different diameters. It can be seen from the left graph of the figure that, at  $T = 30$  K, the line width depends almost linearly on  $x$ . This can be understood as follows. At these temperatures the broadening is determined by a dipolar contribution in addition to an almost constant hyperfine contribution  $H_{HF}$ . Exchange narrowing effects due to nearest-neighbor exchange  $J_{nn}$  are negligible as basically all Mn-ions with Mn nearest neighbors have formed antiferromagnetic pairs. The line width at low temperature can be described as:

$$\Delta H = H_{HF} + H_{dip} \approx H_{HF} + C_{dip} \cdot x. \quad (6)$$

where the dipolar field  $H_{dip} \approx C_{dip}x$  in mean-field approximation [22,23]. The fits in the left graph of Fig. 8 show, that Eq. 6 well describes the observed line width behavior in the  $Zn_{1-x}Mn_xS$  wires, in particular for the 6 and 9 nm samples. The corresponding value of  $H_{HF}$  is about 150 Oe for these samples which is half the extension of the six fine structure peaks in Fig. 3. The value for the 3 nm samples is slightly higher.

The high temperature behavior can be explained semi-quantitatively by considering the effect of the nearest-neighbor interaction  $J_{nn}$ , which determines the line-width behavior at temperatures, where the Mn nearest-neighbor pairs are broken up. For this purpose, the Mn-ions in the nanostructure can be divided into two classes: (i) isolated Mn-ions without Mn nearest neighbors and (ii) Mn-ions with one or more nearest neighbors. Furthermore, as in the case of the discussion of the Curie-Weiss parameter  $\Theta$ , the volume of the nanostructure will be divided into a volume close to the surface  $V_{nn}^s$  and a bulk-like volume  $V_{nn}^b$  (see Eq. 4) and the discussion will be based on a zincblende lattice for simplicity. The probabilities for the occurrence of two classes of Mn-ions as a function of  $x$  at the surface and in the bulk of the nanostructure are given by:

$$p_{is}^{b/s} = (1 - x)^\alpha \quad p_{nn}^{b/s} = 1 - p_{is} = 1 - (1 - x)^\alpha \quad (7)$$

where  $\alpha = z_{nn}^b = 12$  in the bulk and  $\alpha = z_{nn}^s \approx z_{nn}^b/2$  at the surface.

It is assumed that, in both volumes, the line-width contribution  $\Delta H_{is}$  of the isolated Mn-ions is well described by fits of the low temperature line-width data according to Eq. 6. The contribution of the Mn-ions with nearest neighbors is calculated as in Ref. [23], but extended to account for additional broadening due to anisotropic exchange. In both regions it holds:

$$\Delta H_{nn} = \frac{\frac{10}{3}H_{dip,nn}^2 + H_{HF}^2 + (H_{ex,nn}^a)^2}{H_{ex,nn}^i}, \quad (8)$$

where  $H_{dip,nn}$  is the mean dipolar exchange field for a Mn-ion with at least one Mn nearest neighbor. For the bulk-like volume, it is defined as:

$$(H_{dip,nn}^b)^2 \approx \frac{3S(S+1)\mu_B^2 g^2}{4} \left[ \frac{1}{d_{nn}^6} + x^2 \cdot \left( \frac{z_{nn}^b - 1}{d_{nn}^6} + \frac{z_{nnn}^b}{d_{nnn}^6} + \frac{z_{nnnn}^b}{d_{nnnn}^6} \right) \right]. \quad (9)$$

where only the nearest neighbors ( $z_{nn}^b = 12, d_{nn} = a/\sqrt{2}$ ), next nearest neighbors ( $z_{nnn}^b = 6, d_{nnn} = a$ ), and next next-nearest neighbors ( $z_{nnnn}^b = 24, d_{nnnn} = \sqrt{3}/2a$ ) are taken into account. In the surface region ( $H_{dip,nn}^s$ )<sup>2</sup> is obtained by replacing the bulk neighbor numbers  $z^b$  of each shell by the corresponding value  $z^s$ . Again, all the neighbor-numbers  $z^s$  are set to half the bulk value.  $H_{HF}$  is the constant value for the hyperfine broadening determined at low  $T$ , and  $H_{ex,nn}^a$  and  $H_{ex,nn}^i$  are the anisotropic and isotropic nearest-neighbor exchange fields. The latter is calculated according to Anderson and Weiss [23]:

$$H_{ex,nn}^i = 2.83 \frac{J_{nn}}{g\mu_B} \sqrt{S(S+1)}. \quad (10)$$

Using  $|J_{nn}| = 16.1\text{ K}$  for (Zn,Mn)S yields an isotropic nearest-neighbor exchange-field of about 90,000 Oe. The total line width as a function of  $x$  and  $d$  then is:

$$\Delta H(x, d) = V^b \sqrt{(\Delta H_{isp}^b)^2 + (\Delta H_{nn}^b p_{nn}^b)^2} + V^s \sqrt{(\Delta H_{isp}^s)^2 + (\Delta H_{nn}^s p_{nn}^s)^2}. \quad (11)$$

$H_{nn}^a$ , the only free parameter in the calculation, is determined by the line width limit at high  $x$ . The calculations were carried out for all three series. In each case, they were performed for bulk as well as for spherical and wire-like nanostructures. The same value for the anisotropic exchange field  $H_{nn}^a$  was used throughout. The best agreement was obtained for an  $H_{nn}^a = 3000\text{ Oe}$ . This value is about one order of magnitude larger than the dipolar field  $H_{dip,nn}^b$  in reasonable agreement with theoretical findings [24]. Exemplarily, the theoretical curves for the 3 and 6 nm series are plotted in Fig. 8. Again, the agreement between theory and experiment is best when a spherical shape of the  $\text{Zn}_{1-x}\text{Mn}_x\text{S}$  nanoparticles is assumed. In particular, the line width decrease with increasing  $x$  is too rapid when a bulk-like situation is considered. The corresponding slope is reduced by surface effects as the probability for Mn-ions with Mn nearest neighbors is much lower in the surface region  $V_{nn}^s$  than in the bulk-like volume  $V_{nn}^b$  for  $0 < x < 0.2$ .

### 3 Concluding Remarks

Changes of the macroscopic observables (e.g. Curie-Weiss parameter  $\Theta$  and EPR line width  $\Delta H$ ) of the paramagnetic phase of (Zn,Mn)S nanoparticles with sizes below 10 nm due to reduced dimensions are observed. It appears that the microscopic coupling between the Mn-ions (e.g. the exchange constants  $J_{nn}$  and  $J_{nnn}$ ) is not altered to a first approximation. The macroscopic modifications arise mainly due to geometrical restrictions, i.e. the number of neighbors in the various shells around a Mn-ion in the surface region are considerably reduced compared to a Mn-ion in the bulk of the structure. This

effect becomes increasingly important with decreasing lateral dimensions of the nanostructure.

### Acknowledgements

The work is supported by the Deutsche Forschungsgemeinschaft in the framework of the SPP 1072 ‘Semiconductor and Metal Clusters as Building Blocks for organized Structures’.

### References

1. C.T. Kresge, M.E. Leonowicz, W.J. Roth, J.C. Vartulli, J.S. Beck, *Nature* **359**, 710 (1992).
2. D. Zhao, J. Feng, Q. Huo, N. Melosh, G.H. Fredrickson, B.F. Chmelka, and G.D. Stucky, *Science* **279**, 548 (1998).
3. F.J. Brieler, M. Fröba, L. Chen, P.J. Klar, W. Heimbrodt, H.-A. Krug von Nidda, and A. Loidl, *Chem. Eur. J.* **8**, 185 (2002).
4. F. Brieler, P. Grundmann, M. Fröba, L. Chen, P.J. Klar, W. Heimbrodt, H.-A. Krug von Nidda, A. Loidl, and T. Kurz, *J. Am. Chem. Soc.* **126**, 797 (2004).
5. L. Chen, P.J. Klar, W. Heimbrodt, F. Brieler, and M. Fröba, *Appl. Phys. Lett.* **76**, 3531 (2000).
6. W. Xu, Y. Liao, and D.L. Akins, *J. Phys. Chem. B* **106**, 11127 (2002).
7. H. Parala, H. Winkler, M. Kolbe, A. Wohlfahrt, R.A. Fischer, R. Schmehel, and H. von Seggern, *Adv. Mater.* **12**, 1050 (2000).
8. W.-H. Zhang, J.-L. Shi, H.-R. Chen, Z.-L. Hua, and D.S. Yan, *Chem. Mater.* **13**, 684 (2001).
9. L. Chen, H. Falk, P.J. Klar, W. Heimbrodt, F. Brieler, M. Fröba, H.-A. Krug von Nidda, A. Loidl, Z. Chen, and Y. Oka, *phys. stat. sol. (b)* **229**, 31 (2002).
10. L. Chen, P.J. Klar, W. Heimbrodt, F. Brieler, M. Fröba, H.-A. Krug von Nidda, T. Kurz, and A. Loidl, *J. Appl. Phys.* **93**, 1326 (2003).
11. N. Samarth and J.K. Furdyna, *Phys. Rev. B* **37**, 9227 (1988).
12. J. Schneider, S.R. Sircar, and A.Z. Räuber, *Z. Naturforsch. A* **18**, 980 (1963).
13. P.H. Borse, D. Srinivas, R.F. Shinde, S.K. Date, W. Vogel, and S.K. Kulkarni, *Phys. Rev. B* **60**, 8659 (1999).
14. J. Spalek, A. Lewicki, Z. Tarnawski, J.K. Furdyna, R.R. Galazka, and Z. Obuszko, *Phys. Rev. B* **33**, 3407 (1986).
15. W.H. Brumage, C.R. Yarger, and C.C. Lin, *Phys. Rev. A* **133**, A765 (1964).
16. Y.Q. Yang, P.H. Keesom, J.K. Furdyna, and W. Giriat, *J. Solid State Chem.* **49**, 20 (1983).
17. J.K. Furdyna, N. Samarth, R.B. Frankel, and J. Spalek, *Phys. Rev. B* **37**, 3707 (1988).
18. Y. Shapira, S. Foner, D. Heiman, P.A. Wolff, and C.R. McIntyre, *Solid State Commun.* **71**, 355 (1989).
19. R.E. Kremer and J.K. Furdyna, *Phys. Rev. B* **31**, 1 (1985).
20. S.B. Oseroff, *Phys. Rev. B* **25**, 6584 (1982).
21. H.A. Sayad and S.M. Bhagat, *Phys. Rev. B* **31**, 591 (1985).
22. J.H. Van Vleck, *Phys. Rev.* **74**, 1168 (1948).
23. P.W. Anderson and P.R. Weiss, *Rev. Mod. Phys.* **25**, 269 (1953).
24. B.E. Larson and H. Ehrenreich, *Phys. Rev. B* **39**, 1747 (1989).

# Disorder Effects in Diluted Magnetic Semiconductors

Carsten Timm and Felix von Oppen

Institut für Theoretische Physik, Freie Universität Berlin  
Arnimallee 14, 14195 Berlin, Germany

**Abstract.** Diluted magnetic semiconductors (DMS) are promising materials for technological applications as well as interesting from the basic-physics point of view. It has become clear that disorder plays a crucial role in DMS due to the presence of many charged defects and the random positions of impurity spins. In this paper the effect of disorder on transport properties and magnetism in DMS is discussed. It is argued that the impurity positions are correlated due to strong Coulomb interactions between charged defects. These correlations are crucial for the understanding of magnetic properties and the observed metal-insulator transition.

## 1 Introduction

The present interest in diluted magnetic semiconductors (DMS) has been fueled by possible applications in *spintronics* [1]. Particularly promising are ferromagnetic DMS produced by heavily doping technologically relevant semiconductors with transition-metal ions such as manganese. These materials are also fascinating from the basic-physics point of view. They combine strong electronic correlations in the d-shells, unusual transport properties, and strong disorder effects. In this paper we focus on the effects of disorder on transport and magnetism in Mn-doped GaAs. Disorder is expected to be strong due to the presence of a high concentration of charged impurities; the typical distance between these defects is roughly of the same order as the Fermi wave length. See Ref. [2] for a recent review on disorder in DMS.

In  $(\text{Ga},\text{Mn})\text{As}$ , substitutional Mn both acts as an acceptor and provides a localized spin  $S = 5/2$  due to its half-filled d-shell. The doped hole is weakly bound to the acceptor by Coulomb attraction, forming a hydrogenic impurity state split off from the valence band [3,4]. The local impurity spin is coupled to the valence-band holes by exchange interaction. The shallow-impurity description probably applies since states dominated by Mn d-orbitals are far from the Fermi energy [3,4,5].

Because of the low Mn solvability,  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  has to be grown by low-temperature molecular beam epitaxy (MBE) at about  $250^\circ\text{C}$ . Homogeneous growth with up to  $x \approx 0.08$  has been achieved. Ferromagnetic order has first been found by Ohno *et al.* [6]. Recently Curie temperatures above 160 K have been reached due to improved control over the defect concentrations [7,8].

Since Mn is an acceptor,  $(\text{Ga},\text{Mn})\text{As}$  is of p-type. However, the observed hole concentration is lower than the concentration of acceptors due to compensation, probably by arsenic antisites (As at cation sites) and Mn interstitials. Both types of defects are *double donors*. Antisites are expected for low-temperature growth [9]. A high concentration of Mn is believed to lead to increased incorporation of the oppositely charged antisites [10,11,12]. When  $\text{As}_4$  quadruplets are cracked before they arrive at the surface, the antisite concentration can be strongly reduced [13,7,8]. The presence of Mn interstitials has been proposed in Ref. [14] and demonstrated in channeling Rutherford backscattering experiments by Yu *et al.* [15], where about 17% of Mn impurities have been found in interstitial positions.

The ferromagnetic interaction between the impurity spins is carrier-mediated, as demonstrated by electric field-effect experiments [17,18,19]. On the other hand, in  $(\text{Ga},\text{Mn})\text{As}$  a metal-insulator transition between ferromagnetic *insulating* samples with small Mn concentration  $x$  and ferromagnetic *metallic* samples with larger  $x$  is observed [9,16]. Even for the most metallic samples the resistivity is relatively high, of the order of  $10^{-3}$  to  $10^{-2} \Omega \text{ cm}$  [9,20,13], showing that disorder is rather strong even in this regime.

This paper gives arguments why the defects are present in correlated, not random, positions [11,21,2]. In Sec. 2 Monte Carlo (MC) simulations are used to support this proposition. In Secs. 3 and 4 we show that the experimental results for transport and magnetic properties can only be understood assuming such correlations.

## 2 Defect Positions, Growth, and Annealing

In the present section we consider the spatial distribution of defects, i.e., substitutional Mn, As antisites, and interstitial Mn. We explain why the defect positions should be strongly correlated. We employ a rather simple model, which focuses on the screened Coulomb interaction between charged defects. For small defect separations other contributions to the interaction also become relevant. Ab-initio calculations for defect clusters [22,23,24,25] can be used to gain information on these “chemical” contributions. However, our conclusions are unaffected by the specific form of the short-range interaction.

In heavily compensated DMS the electronic screening of the Coulomb interaction between charged defects is rather weak [26]. Therefore, the electronic screening length is large compared to the typical separation between defects. This should lead to two effects [11,21,2]: Firstly, during growth defects of the same (opposite) charge are predominantly incorporated with larger (smaller) separations. Secondly, upon annealing defects further reduce their Coulomb energy through defect diffusion. There is another important effect relevant during annealing: Mn interstitials are more mobile than the other defects [15] due to lower energy barriers for their motion. Conversely, antisites are believed to be rather immobile at typical annealing temperatures

(250°C) [27]. During annealing interstitial Mn seems to diffuse out of the bulk to the free surface of the DMS film [28]. Experiments show that capping of a (Ga,Mn)As film with GaAs suppresses the annealing effects [28]. This suggests that the annealing dependence of physical quantities of these high-quality samples is dominated by out-diffusion of interstitials and not by redistribution of defects within the bulk.

## 2.1 Defect Clusters

We first consider a simplified model where the compensation is assumed to be completely due to antisites. We start from a random distribution of Mn impurities and antisites on the cation sublattice, where the density of antisites is determined by charge neutrality from the observed hole concentrations. The Hamiltonian reads

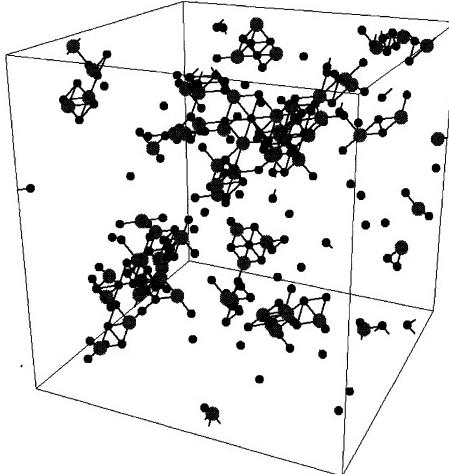
$$H = \frac{1}{2} \sum_{i,j} \frac{q_i q_j}{\epsilon r_{ij}} e^{-r_{ij}/r_{\text{scr}}}, \quad (1)$$

where  $q_i$  are the defect charges and  $r_{ij}$  is their separation. Relative to the cation sublattice, Mn impurities (antisites) carry charge  $q = -1$  (+2). The screening length  $r_{\text{scr}}$  is obtained from nonlinear screening theory [26]; it hardly affects the small-scale defect correlations. We perform MC simulations for the Hamiltonian  $H$  at the growth/annealing temperature 250°C for systems of  $20 \times 20 \times 20$  conventional face-centered-cubic (fcc) unit cells with periodic boundary conditions, unless stated otherwise. The simulation is terminated when the spatial correlation function of the Coulomb potential (discussed below) does not change anymore.

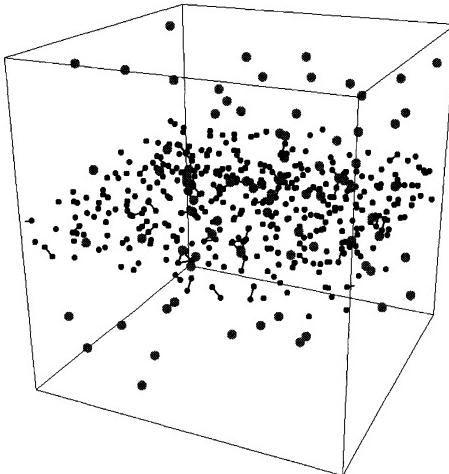
As an illustration, Fig. 1 shows an equilibrated configuration of defects for Mn concentration  $x = 0.05$  and  $p = 0.3$  holes per Mn. Both substitutional Mn and antisites have been assumed to be mobile. This plot is for a supercell size of  $10 \times 10 \times 10$  to make it more clear. It is obvious that the defects form small clusters, where typically an antisite is surrounded by several Mn impurities. The clustering leads to a strongly reduced Coulomb energy.

Before we turn to a quantitative analysis of the MC results, we consider what happens if we do not start from a random configuration. This is motivated by experiments on *digital heterostructures* consisting of half monolayers of MnAs separated by thick layers of GaAs [29]. All samples are insulating for  $T \rightarrow 0$  [29]. Part of the Mn apparently diffuses out of the MnAs layers [29]. We have performed MC simulations for a superlattice of single layers with half the Ga replaced by Mn separated by low-temperature GaAs with 0.25% antisites. The superlattice period is 20 fcc unit cells. Figure 2 shows the configuration when some interdiffusion has occurred.

It is known that a significant fraction of Mn forms interstitials [15]. There are two relevant non-equivalent interstitial positions, namely positions coordinated by four As or by four Ga ions [30]. Ab-initio calculations indicate

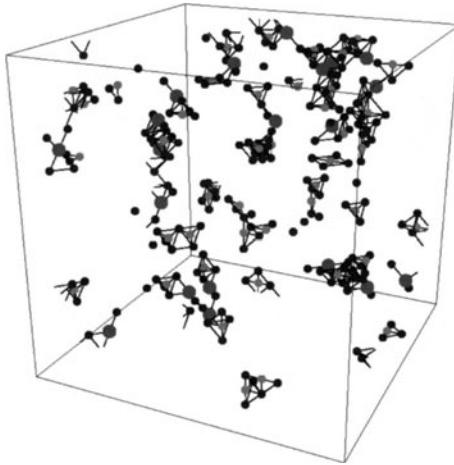


**Fig. 1.** Equilibrated configuration of defects on a  $10 \times 10 \times 10$  lattice for Mn concentration  $x = 0.05$  and  $p = 0.3$  holes per Mn. The host lattice is not shown. Small dark circles denote Mn impurities, whereas large circles are As antisites. Solid bars connect defects on nearest-neighbor sites



**Fig. 2.** Configuration of defects for a superlattice of half monolayers of MnAs between thick layers of low-temperature GaAs. A supercell of  $20 \times 20 \times 20$  fcc unit cells has been used. 2000 MC steps have been performed, leading to interdiffusion of Mn into the GaAs layers. The configuration is still far from equilibrium

that the As-coordinated T(As<sub>4</sub>) position is lower in energy by 0.3 to 0.35 eV compared to the Ga-coordinated position [8,31]. We have performed MC simulations including charged ( $q = +2$ ) interstitials and assuming the onsite energy in the T(As<sub>4</sub>) position to be 0.3 eV lower [2]. Figure 3 shows an equilibrated configuration for a concentration of *substitutional* Mn of  $x = 0.05$  as above, but concentrations of 0.005 of antisites and 0.0125 of Mn interstitials. This corresponds to the same hole concentration as in Fig. 1. All defects have been assumed to be mobile. Again, the defects form small clusters.



**Fig. 3.** Equilibrated configuration of defects on a  $10 \times 10 \times 10$  lattice for substitutional Mn concentration  $x = 0.05$  and  $p = 0.3$  holes per substitutional Mn. This requires a concentration of 0.0175 of compensating double donors, which are assumed to be antisites with a concentration of 0.005 and Mn interstitials (small light circles) with a concentration of 0.0125

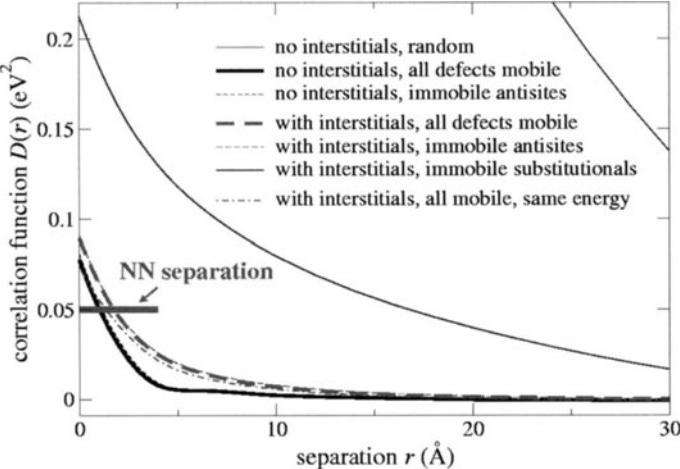
## 2.2 Disorder Potential

The cluster formation can be discussed quantitatively by introducing the correlation function  $D(r)$  of the disorder potential  $V(\mathbf{r})$ :

$$D(r) \equiv \langle V(\mathbf{r}) V(\mathbf{r}') \rangle_{|\mathbf{r} - \mathbf{r}'| = r} - \langle V \rangle^2. \quad (2)$$

$V(\mathbf{r})$  is a sum over screened Coulomb potentials from all impurities. This potential determines the localization properties of the valence-band holes. Obviously,  $\Delta V \equiv \sqrt{D(0)}$  is the *width* of the distribution of  $V(\mathbf{r})$ . Figure 4 shows  $D(r)$  for various assumptions on the type and mobility of defects. The equilibration strongly reduces  $D(r)$  and  $\Delta V$  in all cases. However,  $\Delta V$  is still *not* small compared to the Fermi energy so that disorder cannot be neglected [11,2]. Without interstitials (cf. Fig. 1) the disorder potential becomes short-range correlated. In equilibrium  $D(r)$  decays on the scale of the nearest-neighbor separation due to the screening of the *compensated* Mn impurities [21,2]. The remaining uncompensated Mn ions cannot be screened by antisites and their contribution decays on a longer length scale determined by their density. Assuming antisites to be immobile [27] has practically no effect.

The interstitials do not change the qualitative picture.  $D(r)$  still decays on the length scale of the nearest-neighbor separation. This result changes only weakly if we assume the two different interstitial sites to have the same energy since the energy difference of 0.3 eV is unimportant compared to the Coulomb energy. Again, it makes no difference if the antisites are treated as immobile. The short-range correlations of  $V(\mathbf{r})$  simplify transport calculations [32]. However, if we also assume the *substitutional* Mn to be immobile, the absolute value and the width of the correlation function drastically increase. In this case the small concentration of Mn interstitials cannot effectively screen the Mn substitutionals.



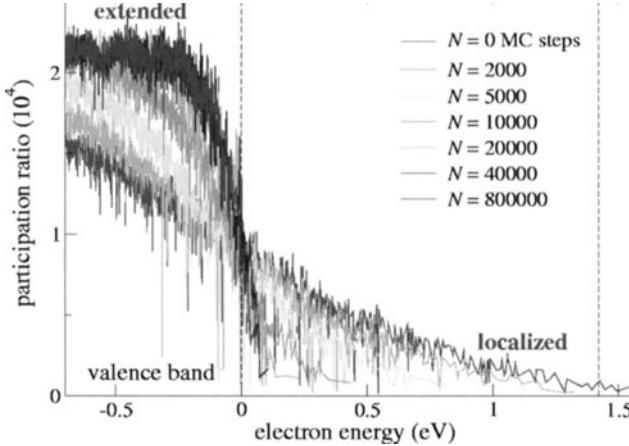
**Fig. 4.** Correlation function  $D(r)$  of the disorder potential for a  $20 \times 20 \times 20$  supercell. A concentration  $x = 0.05$  of substitutional Mn and  $p = 0.3$  holes per substitutional Mn have been assumed. The thin solid line lying mostly outside the scope of the figure shows the result for a random distribution of defects (in this case without interstitials). The other curves show  $D(r)$  for equilibrated configurations with or without interstitials under various assumptions on the mobility of different defect species as given in the legend. The heavy horizontal bar denotes the nearest-neighbor separation on the cation sublattice

The scenario of immobile substitutionals agrees with the suggestion that annealing mostly leads to outdiffusion of interstitials and not to redistribution of substitutional defects [28]. In this case defect correlations are expected to develop mainly during growth. We argue in the following that they have to be present to explain the observations. On the other hand, the experiments on digital heterostructures [29] would be hard to interpret if substitutional Mn were totally immobile. The observed interdiffusion of Mn would require a probably unrealistically high fraction of interstitial Mn. The presence of cation vacancies could make substitutional defects more mobile [20].

We note that in Ref. [25] small clusters of three Mn defects are studied within density-functional theory. Clusters made up of two substitutional and one interstitial defects are found to be favored due to their Coulomb attraction, in agreement with the above discussion. As noted above, ab-initio theory captures additional contributions to the energy not included here.

### 3 Hole States

Next, we consider the properties of the valence-band holes subjected to the impurity potential  $V(\mathbf{r})$ . In particular, we are interested in their spectrum and localization properties. To exhibit the main physics we again employ



**Fig. 5.** Participation ratio as a function of energy for  $x = 0.05$  and  $p = 0.3$  after various numbers  $N$  of MC steps increasing from the flattest to the steepest curve. The unperturbed band edges are marked by vertical dashed lines

a simple model, which consists of the envelope-function and parabolic-band approximations and starts from the single-hole Hamiltonian

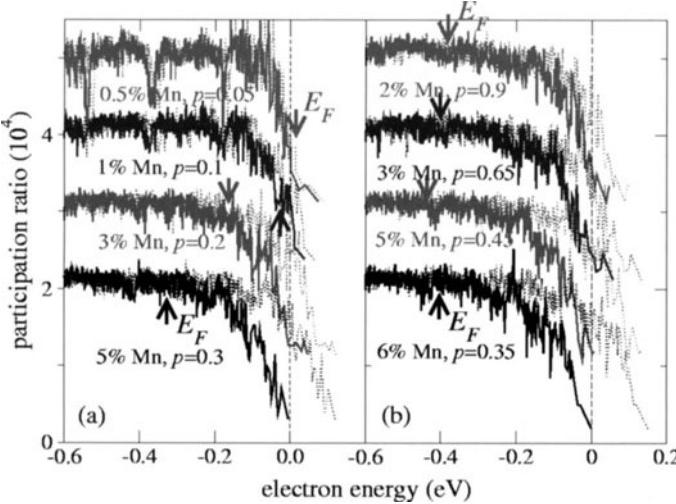
$$H = -\frac{\hbar^2}{2m^*} \nabla^2 + V(\mathbf{r}). \quad (3)$$

The hole Hamiltonian diagonalized numerically in a plane-wave basis [11]. The calculations are done for spin-less holes, since the additional disorder introduced by the exchange interaction is found to be small. We reintroduce the exchange below. The diagonalization yields the energy spectrum and eigenfunctions  $\psi_n(\mathbf{r})$ , from which we calculate the participation ratios

$$\text{PR}(n) = \frac{\left[ \sum_{\mathbf{r}} |\psi_n(\mathbf{r})|^2 \right]^2}{\sum_{\mathbf{r}} |\psi_n(\mathbf{r})|^4}. \quad (4)$$

The PR allows to estimate the position of the mobility edge in the valence band since it scales with system size for extended states but essentially remains constant for localized states.

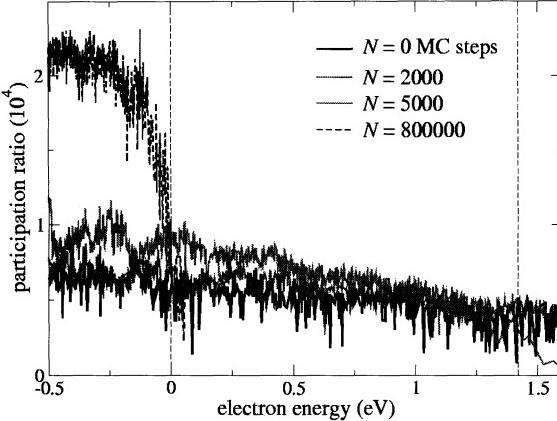
Figure 5 shows the PR as a function of energy for a Mn concentration of  $x = 0.05$  and  $p = 0.3$  holes per Mn [9], for various numbers of MC steps,  $N$  [2]. For fully random defects ( $N = 0$ ) the valence-band edge is smeared by disorder to such an extent that the gap becomes filled, in contradiction to experiments. On the other hand, after equilibration only very few states remain in the gap. Thus the defect clustering is *required* to explain the persistence of the gap. From finite-size scaling results (not shown) we know that the PR in the flat region and the upper part of the slope scales with system size, whereas it becomes independent of system size in the band tail [11,21].



**Fig. 6.** Participation ratio for equilibrated configurations as a function of energy for various parameters  $x$  and  $p$  given in the plot, (a) for the (Ga,Mn)As samples of [9], (b) for the samples of [13]. For the exchange splitting of the band full magnetization of Mn spins has been assumed. The Fermi energy is in each case indicated by an arrow

The mobility edge thus lies on the slope in Fig. 5. Due to numerical constraints the finite-size scaling could not be extended to large enough systems to pinpoint the mobility edge exactly.

Transport properties are dominated by the states close to the Fermi energy  $E_F$ . To find  $E_F$  one has to take the splitting of the valence band by the exchange interaction into account. As noted above, the disorder due to this interaction is small. However, its *average* is not negligible. Figure 6 shows the PR as a function of energy for parameters appropriate for (a) the samples measured by Matsukura, Ohno *et al.* [9] and (b) the samples of Edmonds *et al.* [13]. The defect concentration has relatively little effect on the PR curve due to the strong ionic screening. In (a) the main effect is the shift of the Fermi energy due to the strongly changing hole concentration. The states at the Fermi energy are extended for  $x \geq 0.03$ , while they are localized for  $x = 0.005$  [11,2], in reasonable agreement with the experimentally observed metal-insulator transition [9]. On the other hand, in (b) the Fermi energy hardly changes at all and stays well in the extended region, consistent with the experimental observation of metallic transport from  $x = 0.0165$  to  $x = 0.088$  in these samples [13,33]. The growth technique of Ref. [13] leads to very low compensation at small Mn concentrations and to *increasing* compensation with increasing Mn doping, oppositely to what is observed in Refs. [9].



**Fig. 7.** Participation ratio as a function of electron energy for a superlattice of half monolayers of MnAs between thick layers of low-temperature GaAs, see Fig. 2, for various numbers of MC steps

For fully random defects the states at the Fermi energy would in all cases show a strong tendency towards localization. Thus cluster formation is required to understand why (Ga,Mn)As becomes metallic at all.

Finally, we show in Fig. 7 the PR for the simulation starting from a half layer of MnAs between thick layers of low-temperature GaAs, see Fig. 2. The broadening of the band due to disorder is very strong for weak interdiffusion ( $N = 2000$ ) and the PR is strongly reduced. This strong tendency towards localization is in agreement with experiments [29]. Ionic screening is rather inefficient since the Mn defects and the oppositely charged antisites first have to come close together to form clusters. The PR curve for  $N = 800000$  shows that eventually states close to equilibrium are reached.

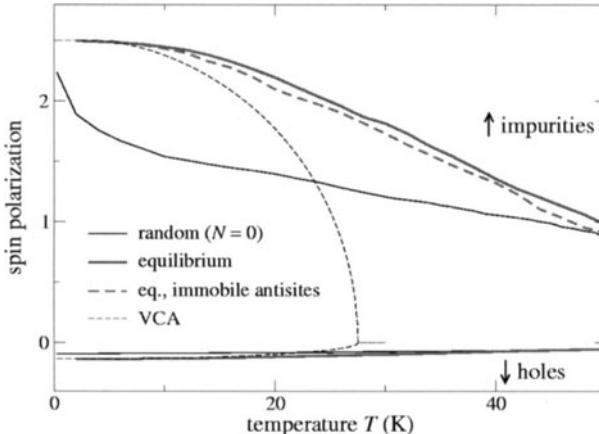
## 4 Magnetic Order

So far, we have only included the exchange interaction with the Mn impurities through a homogeneous exchange field, i.e., at the level of the virtual crystal approximation (VCA) [34]. We now study the magnetic order within a selfconsistent mean-field approximation, which, importantly, fully includes the disorder [11,2]. We start from a Zener model with the Hamiltonian

$$H = \sum_{n\sigma} \xi_n c_{n\sigma}^\dagger c_{n\sigma} - J_{pd} \sum_i \mathbf{s}_i \cdot \mathbf{S}_i, \quad (5)$$

where the  $\mathbf{S}_i$  are Mn spins ( $S = 5/2$ ) and

$$\mathbf{s}_i \equiv \sum_{n\sigma n'\sigma'} c_{n\sigma}^\dagger \psi_n^*(\mathbf{R}_i) \frac{\sigma\sigma'}{2} \psi_{n'}(\mathbf{R}_i) c_{n'\sigma'} \quad (6)$$



**Fig. 8.** Spin polarization of Mn spins (up) and hole spins (down) as functions of temperature for  $x = 0.05$  and  $p = 0.3$  for random and equilibrated/clustered configurations. The spin polarization for an equilibrated system with immobile antisites is also shown. For comparison, the curves labeled by ‘‘VCA’’ show the magnetizations obtained from a theory that totally neglects disorder

are hole-spin polarizations at the Mn sites.  $\xi_n$  and  $\psi_n$  are the hole eigenenergies and eigenfunctions, respectively, for vanishing magnetic interaction obtained in the previous section. Note that  $H$  contains the full Coulomb disorder potential.  $\xi_n$  includes the chemical potential. The antiferromagnetic exchange interaction is assumed to be local. In a more realistic model one should take its nonzero range into account, which stems from the hybridization between Mn d-orbitals and As p-orbitals. The finite range is important for the anisotropy of the effective exchange interaction between impurity spins [35].

The Hamiltonian (5) is decoupled in a mean-field approximation without taking a spatial average [11,21,2]. At each impurity site the product  $\mathbf{s}_i \cdot \mathbf{S}_i$  is replaced by  $\langle \mathbf{s}_i \rangle \cdot \mathbf{S}_i + \mathbf{s}_i \cdot \langle \mathbf{S}_i \rangle - \langle \mathbf{s}_i \rangle \cdot \langle \mathbf{S}_i \rangle$ . The averaged hole spins  $\langle \mathbf{s}_i \rangle$  are expressed in terms of the mean-field Zeeman splittings of the hole states  $n$ . This leads to a large set of coupled mean-field equations which are solved numerically. A collinear magnetization is assumed. One can show that the equation for the Curie temperature  $T_c$ , obtained by linearizing the mean-field equations, is unchanged by dropping this assumption.

Figure 8 shows the magnetization curves for Mn and hole spins for  $x = 0.05$  and  $p = 0.3$  for random and equilibrium defect configurations, assuming compensation by antisites [11,21,2]. Since these results are obtained from mean-field theory, they are most reliable at low temperatures. However, it has been suggested that mean-field theory is quite good for metallic (Ga,Mn)As in the whole temperature range [36,37,34]. The Mn magnetization curve shows *upwards* curvature for random defects. Similar magnetization curves have been calculated by many different approaches, see Ref. [2]. This general shape

seems to be a robust feature of highly disordered ferromagnets. Within mean-field theory, the anomalous shape is due to the localization tendency of the holes, which leads to a shorter-range effective Mn-Mn spin interaction and thus to a broader distribution of effective fields acting on these spins.

For the clustered configuration the magnetization curve shows *downward* curvature and is more Brillouin-function-like at low temperatures. The crossover from upward to downward curvature agrees qualitatively with experimental annealing studies [20]. Conversely, the assumption of clustered defects is crucial to understand the downward curvature observed for optimally annealed samples. Figure 8 also shows that assuming immobile antisites again does not have a strong effect. Neglecting the disorder by setting  $V(\mathbf{r}) = 0$  and replacing the Mn spins by a homogeneous spin density (VCA) leads to a strongly reduced Curie temperature. This may be an artifact of mean-field theory, but since the impurity-spin polarization exceeds the VCA result even at the lowest temperatures, where mean-field theory should be valid, we suggest that disorder indeed has the tendency to stabilize ferromagnetism.

We do not expect Mn interstitials to change this picture qualitatively. Manganese impurities in adjacent substitutional and interstitial positions should interact antiferromagnetically due to superexchange. The strength of this interaction is at least  $-26$  meV [30], leading to the formation of spin singlets that do not participate in ferromagnetic order. This suggests to take only the substitutionals into account that are not paired up in singlets.

## 5 Summary

Correlated defect clusters develop naturally in DMS due to the strong Coulomb interactions between oppositely charged defects. The clustering is expected to take place during growth or annealing and leads to a strongly reduced impurity potential. This effect is necessary to explain the experimental observations regarding the metal-insulator transition as a function of Mn doping and the persistence of the semiconductor gap. We also find that the magnetization curve crosses over from an anomalous shape to a more Brillouin-function-like dependence, in agreement with experiments. In high-quality samples the cluster formation has to take place already during growth since they already show metallic behavior and convex magnetization curves *as grown* [28]. While the simple model employed in this study precludes quantitative comparison with experiments, the general conclusions do not depend on model details.

## Acknowledgements

Part of the results have been obtained in collaboration with F. Höfling and F. Schäfer. We thank T. Dietl, S.C. Erwin, G.A. Fiete, P. Kacman, A.H. MacDonald, M.E. Raikh, J. Sinova, and G. Zaránd for helpful discussions.

## References

1. S.A. Wolf *et al.* *Science* **294**, 1488 (2001).
2. C. Timm, *J. Phys.: Condensed Matter* **15**, R1865 (2003).
3. M. Linnarsson *et al.*, *Phys. Rev. B* **55**, 6938 (1997).
4. J. Okayabashi *et al.*, *Phys. Rev. B* **58**, R4211 (1998).
5. T. Dietl, *Semicond. Sci. Technol.* **17**, 377 (2002).
6. H. Ohno *et al.*, *Appl. Phys. Lett.* **69**, 363 (1996).
7. K.C. Ku *et al.*, *Appl. Phys. Lett.* **82**, 2302 (2003).
8. K.W. Edmonds *et al.*, *Phys. Rev. Lett.* **92**, 037201 (2004).
9. F. Matsukura, H. Ohno, A. Shen, Y. Sugawara, *Phys. Rev. B* **57**, R2037 (1998); H. Ohno, *Science* **281**, 951 (1998); H. Ohno, *J. Magn. Magn. Mat.* **200**, 110 (1999).
10. B. Grandidier *et al.*, *Appl. Phys. Lett.* **77**, 4001 (2000).
11. C. Timm, F. Schäfer, F. von Oppen, *Phys. Rev. Lett.* **89**, 137201 (2002).
12. L. Bergqvist *et al.*, *Phys. Rev. B* **67**, 205201 (2003).
13. K.W. Edmonds *et al.*, *Appl. Phys. Lett.* **81**, 3010 (2002).
14. J. Mašek, F. Máca, *Acta Phys. Pol. A* **100**, 319 (2001).
15. K.M. Yu *et al.*, *Phys. Rev. B* **65**, 201303(R) (2002).
16. H. Ohno, F. Matsukura, *Sol. State Commun.* **117**, 179 (2001).
17. H. Ohno *et al.*, *Nature* **408**, 944 (2000).
18. D. Chiba, M. Yamanouchi, F. Matsukura, H. Ohno, *Science* **301**, 943 (2003).
19. A.M. Nazmul, S. Kobayashi, S. Sugahara, M. Tanaka, cond-mat/0309532 (2003).
20. S.J. Potashnik *et al.*, *Appl. Phys. Lett.* **79**, 1495 (2001).
21. C. Timm, F. von Oppen, *J. Supercond.* **16**, 23 (2003).
22. M. van Schilfgaarde, O.N. Mryasov, *Phys. Rev. B* **63**, 233205 (2001).
23. H. Raebiger, A. Ayuela, R.M. Nieminen, cond-mat/0307364 (2003).
24. J.M. Sullivan *et al.*, *Phys. Rev. B* **68**, 235324 (2003).
25. P. Mahadevan, A. Zunger, cond-mat/0309502 (2003).
26. B.I. Shklovskii, A.L. Efros, *Electronic Properties of Doped Semiconductors* (Springer-Verlag, Berlin 1984).
27. M. Suezawa, *Int. Conf. on Science and Technology of Defect Control in Semiconductors* (Yokohama, Japan), vol. II, K. Sumino (Ed.) (North Holland, Amsterdam 1990), p. 1043.
28. M.B. Stone *et al.*, *Appl. Phys. Lett.* **83**, 4568 (2003).
29. R.K. Kawakami *et al.*, *Appl. Phys. Lett.* **77**, 2379 (2000); E. Johnston-Halperin *et al.*, *Phys. Rev. B* **68**, 165328 (2003).
30. J. Mašek, F. Máca, cond-mat/0308568 (2003).
31. J.M. Sullivan, S.C. Erwin, unpublished (2003).
32. C. Timm, F. von Oppen, F. Höfling, *Phys. Rev. B* (in press), cond-mat/0309547 (2003).
33. B.L. Gallagher, private communication (2003).
34. J. König, J. Schliemann, T. Jungwirth, A.H. MacDonald, *Electronic Structure and Magnetism of Complex Materials, Springer Series in Material Sciences*, vol. **54**, D.J. Singh and D.A. Papaconstantopoulos (Eds.) (Springer-Verlag, Berlin 2003), pp. 163.
35. C. Timm, A.H. MacDonald, unpublished (2004).
36. T. Dietl *et al.*, *Science* **287**, 1019 (2000).
37. T. Jungwirth *et al.*, *Appl. Phys. Lett.* **81**, 4029 (2002); T. Jungwirth *et al.*, *Appl. Phys. Lett.* **83**, 320 (2003).

# Magnetic Anisotropy in Ferromagnetic III-Mn-V Semiconductors: Issues and Observations

J. K. Furdyna<sup>1</sup>, X. Liu<sup>1</sup>, T. Wojtowicz<sup>1,2</sup>, W. L. Lim<sup>1</sup>, U. Welp<sup>3</sup>, and V. K. Vlasko-Vlasov<sup>3</sup>

<sup>1</sup> Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA

<sup>2</sup> Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland

<sup>3</sup> Materials Science Division, Argonne National Laboratory, Argonne, IL 60439, USA

**Abstract.** Magnetic anisotropy (MA) in III-V-based ferromagnetic (FM) semiconductor alloys (e.g.,  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ ) is of critical importance for the design and implementation of spin-dependent devices based on these materials. This property can be “engineered” and controlled either by imposing appropriate strain conditions on the FM film, or by choosing the hole concentration  $p$ . Here it is especially important that the choice of  $p$  can be controlled by external mechanisms, such as illumination or voltage. In this paper we present the information on MA in FM semiconductors obtained through a series of complementary methods: ferromagnetic resonance, planar Hall effect, direct imaging of FM domains by magneto-optical methods, and SQUID magnetization data. Special attention is given to the recently-observed re-orientation of the easy axis in several of the  $\text{III}_{1-x}\text{Mn}_x\text{V}$  alloys as a function of temperature and/or illumination, and to the somewhat surprising uniaxial anisotropy in the basal (001) plane which some of these FM films exhibit, presumably as a consequence of surface reconstruction during the growth.

## 1 Introduction

Recently much effort has been devoted to the growth and characterization of ferromagnetic (FM) semiconductors materials, with an eye on future “spintronic” applications [1,2]. In particular, much progress has been made in the fabrication of III-Mn-V alloys (for which the Curie temperatures  $T_C$  exceeding 150 K have been reported [3,4]), and in their materials engineering [5,6]. It is especially important that magnetic properties of the FM heterostructure  $\text{In}_{1-x}\text{Mn}_x\text{As}/\text{GaSb}$  can be controlled by external perturbations, such as illumination [7,8] or an applied electric field [9,10]. One of key properties of these FM materials is magnetic anisotropy (MA), since it is the basic parameter for developing future spin-based devices (e.g., spin-injectors [11], tunnel junctions [12], devices based on the giant planar Hall effect (GPHE) [13], and nano-constrictions exhibiting giant magneto-resistance [14]).

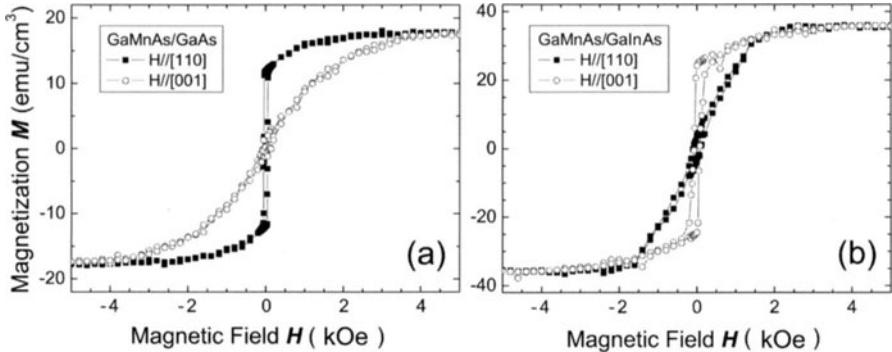
It is now well established that the ferromagnetic interaction in the III-Mn-V materials is mediated by free holes [15], and that MA in FM III-Mn-V

originates from the anisotropic character of the carriers mediating the FM exchange between the localized spins of the (isotropic) Mn<sup>++</sup> ions [16,17]. It has been amply demonstrated that the MA of Ga<sub>1-x</sub>Mn<sub>x</sub>As films is largely controlled by epitaxial strain, the compressive and tensile strain inducing an in-plane and an out-of-plane orientation of the magnetic moments, respectively [18]. The above issues have already led to a broad range of experimental studies. [19,20,21] In this paper we review the information on MA in FM semiconductors obtained through a series of complementary methods: ferromagnetic resonance (FMR) [21], planar Hall effect (PHE), direct imaging of FM domains by magneto-optical methods [22], and SQUID magnetization measurements. Special attention will be given to the manipulation of MA in III-Mn-V semiconductors through low-temperature annealing and modulation doping; to the recently-observed re-orientation of the easy axis in several of the III<sub>1-x</sub>Mn<sub>x</sub>V alloys as a function of temperature and/or illumination [8]; and to the somewhat surprising uniaxial anisotropy in the basal (001) plane of the FM layers which some of these cubic materials exhibit [23,24], presumably as a consequence of surface reconstruction during their growth along the [001] direction.

## 2 Uniaxial and Cubic Magnetic Anisotropy

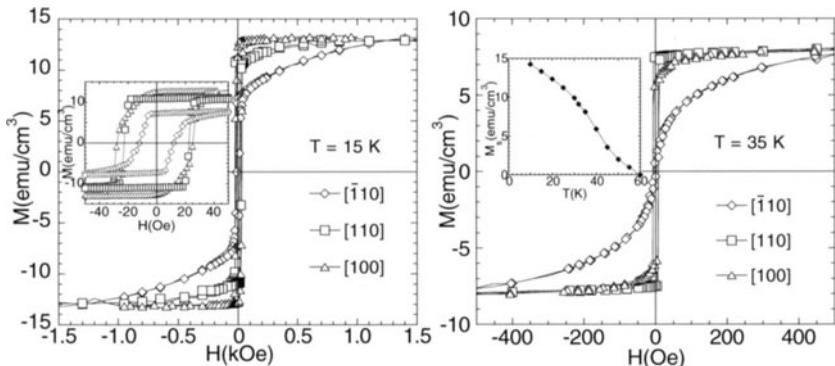
Because of the low solubility of Mn in bulk III-V semiconductors, III<sub>1-x</sub>Mn<sub>x</sub>V alloys with  $x$  sufficiently high to produce cooperative magnetic effects ( $x > 0.01$ ) can only be achieved by low temperature molecular beam epitaxy (LT-MBE), and thus the lattices of these materials are locked to those of their substrates [2]. For examples, x-ray diffraction (XRD) experiments have revealed that the Ga<sub>1-x</sub>Mn<sub>x</sub>As is coherently strained throughout its thickness by the underlying layers (typically GaAs or GaInAs), even for films as thick as 6.8  $\mu\text{m}$  [25]. Based on mean-field theory calculations, it has been shown that for the [001] growth direction the strain splits the heavy hole and light hole bands, producing in turn a uniaxial contribution to the magnetic anisotropy of the III-Mn-V films as follows: the orientation of the easy axis along the growth direction is favored when the strain shifts the heavy holes further down relative to the light holes (i.e., *below* the top of the light-hole sub-band), and an in-plane easy axis occurs in the opposite circumstance [15,16].

Experimentally, it has also been clearly confirmed that a III-Mn-V film can be magnetized more easily along certain crystallographic directions than others. For example, as presented in Fig. 1, dc SQUID magnetization measurements carried out on a 300 nm Ga<sub>0.97</sub>Mn<sub>0.03</sub>As film show two remarkable features, of interest for our discussion of MA. First, these data demonstrate the striking difference in uniaxial anisotropy of GaMnAs under compressive (GaMnAs/GaAs) and tensile strain (GaMnAs/Ga<sub>0.85</sub>In<sub>0.15</sub>As). The hysteresis loops for GaMnAs/GaAs (top panel) clearly show that the easy axis of magnetization is in the plane of the sample. In contrast, in the case of the



**Fig. 1.** Magnetization  $M$  as function of applied magnetic field  $H$  for a  $\text{Ga}_{0.97}\text{Mn}_{0.03}\text{As}$  thin film, measured at  $T = 5\text{ K}$  by SQUID for GaMnAs/GaAs (left panel) and GaMnAs/GaInAs (right panel). The magnetic field is applied either in the [110] or in the [001] direction. GaMnAs grown directly on a (001) GaAs substrate (left panel) is under compressive strain in the layer plane; GaMnAs grown on a  $\text{Ga}_{0.85}\text{In}_{0.15}\text{As}$  buffer (right panel) is under tensile strain. Note that the easy axis is different for the two samples.

GaMnAs/GaInAs sample, which is under tensile strain, the easy axis is obviously normal to the layer plane, as evidenced by the sharp hysteresis loop observed when  $\mathbf{H}$  is applied parallel to the [001] direction (bottom panel). The second conspicuous feature in Fig. 1 involves the hysteresis loops observed when  $\mathbf{H}$  is applied along the hard axes ( $\mathbf{H}$  normal to the plane for GaMnAs/GaAs, and  $\mathbf{H}$  in the plane of the film for the GaMnAs/GaInAs system). We note that in both cases the saturation magnetizations for the easy and hard axes of magnetization coincide at fields of about 3000 G or less, indicating that fields of that magnitude are sufficient to orient all spins along the hard axis (i.e., that they are comparable to the uniaxial anisotropy fields,  $2K_{2\perp}/M$ ). Obviously, in the absence of strain, the III-Mn-V would intrinsically have a purely cubic MA arising from its zinc-blende symmetry. However, the cubic anisotropic character of these materials is usually masked by the effects induced by the aforementioned strain, so that the cubic anisotropy terms can most readily be identified by studying the (001) plane, as demonstrated in Fig. 2. Figures 2a and 2b show magnetization curves for a  $\text{Ga}_{0.97}\text{Mn}_{0.03}\text{As}$  film grown on GaAs (100) substrate at 15 and 35 K for fields along the in-plane [110], [110], and [100] directions. Distinct in-plane anisotropy of the magnetization is clearly revealed. At 15 K, [100] is the easy axis, with essentially the full moment at remanence. Nevertheless, magnetization reversal starts gradually close to  $H = 0$  Oe (see inset in Fig. 2a), followed by sharp switching at the coercive field of about 28 Oe. Hard-axis-like behavior is seen for the [110] direction, with saturation occurring around 1.5 kOe. The magnetization curves for the [100] and [010] directions are identical. The distinct difference in behavior for the [110] and [110] directions is not expected



**Fig. 2.** Magnetization curves for a  $\text{Ga}_{0.97}\text{Mn}_{0.03}\text{As}/\text{GaAs}$  sample at 15 K (a) and 35 K (b) along different crystallographic directions. Inset (a): expanded presentation of the data at 15 K; inset (b): temperature dependence of the saturation magnetization  $M_s$ .

on the grounds of crystal symmetry, and will be discussed in detail below. At 35 K the easy axis has switched to the  $[110]$  direction, whereas the  $\langle\bar{1}10\rangle$  direction continues to display typical hard-axis behavior, with essentially zero remanence. The temperature dependence of the saturation magnetization  $M_s$  is shown in the inset of Fig. 2b.

Further investigation has been carried out in a series of as-grown  $\text{GaMnAs}/\text{GaAs}$  films with thicknesses ranging from 0.2 to 6.8  $\mu\text{m}$ .[25] For all samples we find a clear difference between the  $\langle\bar{1}10\rangle$  and  $\langle110\rangle$  directions, i.e., a uniaxial anisotropy contribution along the  $\langle110\rangle$  direction [23]. It is important to note that the degree of the uniaxial anisotropy is essentially independent of the thickness of the film, which rules out the possibility that the observed uniaxial anisotropy arises from the contribution of the surface layer alone due to the well-known surface reconstruction. Instead, these new results suggest that the uniaxial anisotropy – while it is probably seeded by the anisotropy of the reconstructed initial  $\text{GaAs}(001)$  substrate surface (similar to that seen in  $\alpha\text{-Fe}$  (bcc) films grown on  $\text{GaAs}$  [26]) – its thickness-independence is very likely due to replicating that seeded anisotropy, layer by layer, as the growth continues, thus penetrating the entire  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  film.

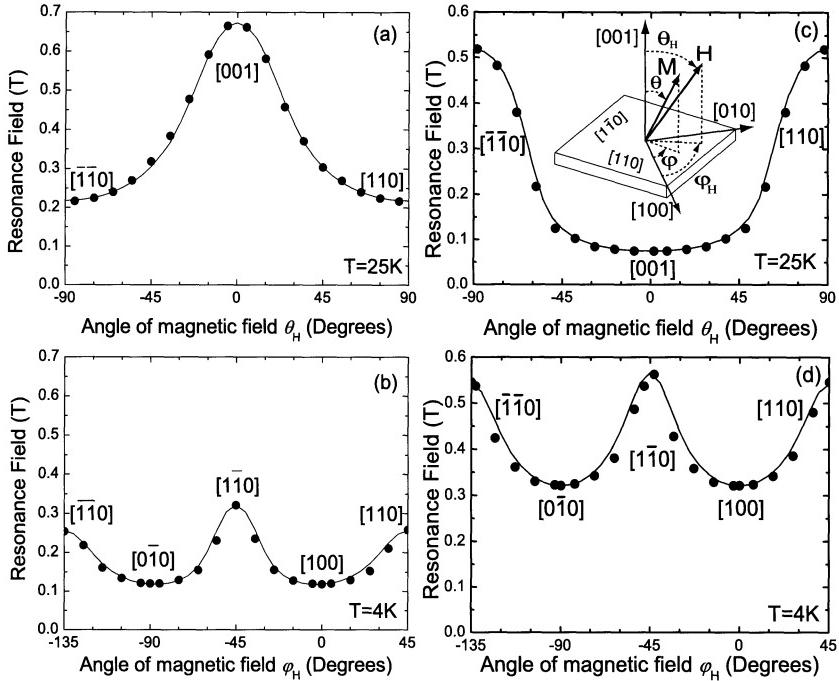
### 3 Ferromagnetic Resonance in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ : Effects of Magnetic Anisotropy

FMR complements the more conventional magnetic studies of FM films by its ability to directly determine MA parameters [27], and in this section we take advantage of this technique to obtain MA information in  $\text{GaMnAs}$  films

under various strain conditions [21]. We use the well-known Landau-Lifshitz-Gilbert equation [28] for the precession of the magnetization to describe the dynamics of FMR in a GaMnAs film, treating GaMnAs as a zinc-blende crystal under tetragonal distortion. We have mapped out the FMR condition for 300 nm  $\text{Ga}_{0.97}\text{Mn}_{0.03}\text{As}$  films under different strain conditions as a function of magnetic field orientation relative to the crystal axes at a fixed temperature. The compressive strain was attained automatically when GaMnAs was grown directly on a GaAs substrate, while the tensile strain was achieved by growing GaMnAs on a relaxed GaInAs buffer, with sufficient In concentration to make the lattice parameter of the buffer larger than that of the FM film. The observed FMR field positions obtained for the GaMnAs/GaAs sample are shown in Fig. 3a and 3b, and those for GaMnAs/GaInAs are in Fig. 3c and 3d. FMR measurements were carried out in the two geometries (i.e.,  $\mathbf{H}$  in the  $(1\bar{1}0)$  plane; and  $\mathbf{H}$  in the  $(001)$  plane), corresponding to the upper (Figs. 3a and 3c) and lower panels (Figs. 3b and 3d) of Fig. 3, respectively. To facilitate discussion, a polar coordinate system is defined as an inset in Fig. 3c.

Note that the lowest resonance field for the GaMnAs/GaAs sample (Fig. 3a and 3b) is observed when  $\mathbf{H}$  lies in the film plane and is parallel to the easy axis – the  $[100]$  direction ( $\theta_{\mathbf{H}} = 90^\circ$ ,  $\varphi_{\mathbf{H}} = 0^\circ$ ), as defined in the inset in Fig. 3c. Figure 3b, which shows the angular dependence of  $H_R$  for  $\mathbf{H}$  in the  $(001)$  plane, is particularly interesting for two reasons. First, it reveals that the magnetic in-plane anisotropy (which we attribute to the cubic anisotropy energy  $K_{4\parallel}$ ) is quite strong. The origin of this behavior can be ascribed to the physical difference between the  $[100]$  and  $[110]$  orientations in the  $(001)$  plane. Second, it is evident from Fig. 3b that the symmetry of the FMR position is not exactly four-fold in the  $(001)$  plane (as would be expected for a purely tetragonal distortion along the  $[001]$  axis), but that a small difference exists between  $H_R$  for  $\mathbf{H}$  applied along the  $[110]$  and  $[\bar{1}10]$  directions.

In Figs. 3c and 3d we show the full angular dependences of the resonance field  $H_R$  for the  $\text{Ga}_{0.97}\text{Mn}_{0.03}\text{As}/\text{Ga}_{0.85}\text{In}_{0.15}\text{As}$  film measured in two geometries at a fixed temperature. In the context of FMR, the reversal of the sign of the strain leads to a large *positive*  $2K_{2\perp}/M$ , which counteracts the demagnetization field, thus shifting the FMR position downward when the magnetic field is applied normal to the plane, as seen in Fig. 3c. Figure 3d shows that in the GaMnAs/GaInAs combination there is also a much stronger four-fold in-plane anisotropy energy  $K_{4\parallel}$ , and again a very small but unmistakable planar two-fold uniaxial anisotropy. It is also interesting to note that – as can be inferred from the non-sinusoidal nature of the angular dependence of  $H_R$  in Fig. 3d – when  $\mathbf{M}$  is in the sample plane, it tends to seek the  $[100]$  direction even when that is not the real easy axis of the tetragonally-strained material as a whole. The solid curves in Fig. 3 represent the positions of  $H_R$  calculated theoretically by using  $g = 2.0$  for the  $\text{Mn}^{++}$  ions and the optimized values



**Fig. 3.** Angular dependence of FMR fields for a  $\text{Ga}_{0.97}\text{Mn}_{0.03}\text{As}/\text{GaAs}$  sample (Fig. 3a and 3b) and  $\text{Ga}_{0.97}\text{Mn}_{0.03}\text{As}/\text{Ga}_{0.85}\text{In}_{0.15}\text{As}$  sample (Fig. 3c and 3d). Fig. 3a and 3c: dc magnetic field  $\mathbf{H}$  and magnetization  $\mathbf{M}$  in the  $(\bar{1}\bar{1}0)$  plane; and Fig. 3b and 3d:  $\mathbf{H}$  and  $\mathbf{M}$  in the  $(001)$  plane. The solid curves in the figure represent the positions of  $H_R$  calculated theoretically. The coordinate system used in figure is shown as an inset in Fig. 3c. The orientation of the dc magnetic field  $\mathbf{H}$  (described by  $\theta_H$  and  $\varphi_H$ ) can be varied continuously in both the  $(\bar{1}\bar{1}0)$  and the  $(001)$  planes. The resulting equilibrium orientations of the magnetization  $\mathbf{M}$  are given by  $(\theta, \varphi)$ .

of uniaxial and cubic anisotropy fields  $2K_i/M$ . The ability to reproduce the observed data is a clear indication that FMR in GaMnAs indeed obeys the model for zinc-blende symmetry with some degree of distortion, induced by the strain and by the in-plane anisotropy along the  $<110>$  directions.

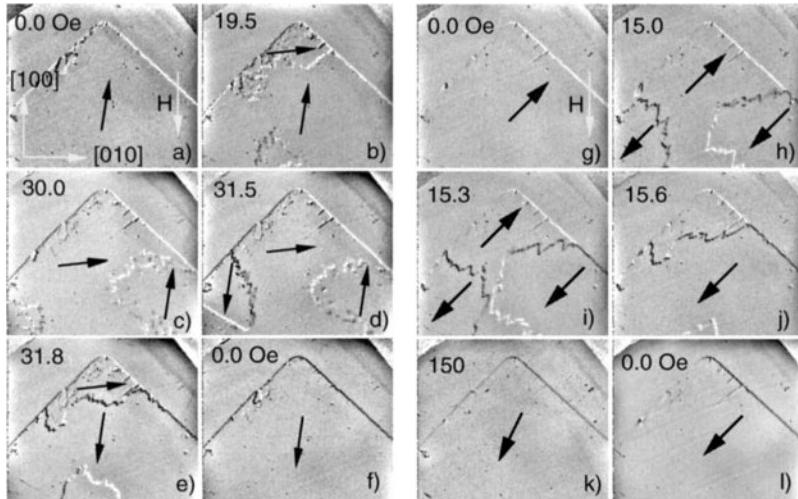
The measurements of FMR show unambiguously that MA plays a decisive role in determining the resonance spectrum of epitaxially grown GaMnAs. Since the crystal structure of epitaxial GaMnAs is slightly tetragonal due to the distortion of its relaxed cubic structure by the lattice mismatch with the substrate, it is expected that the MA of this material will be dominated by uniaxial anisotropy. We have observed uniaxial anisotropy field  $2K_{2\perp}/M$  as large as 4000 G, and larger values can probably be induced by greater lattice mismatches than those used in growing the present samples. The sign of the  $2K_{2\perp}/M$  term is determined by whether the strain in the layer plane is tensile or compressive, as clearly revealed by the positions of FMR observed for GaMnAs/GaAs and GaMnAs/GaInAs, respectively. The cubic (four-fold)

anisotropy fields are also very large in GaMnAs (of the order of 2000 G), as evidenced by the angular variation of FMR observed when the orientation of  $\mathbf{H}$  is varied in the basal (001) plane. Finally, these experiments reveal that there is a small but finite in-plane *uniaxial* anisotropy field  $2K_{2\parallel}/M$ , associated with the slight physical difference between the [110] and [−110] directions of the zinc-blende structure on which the GaMnAs film is grown.

#### 4 Ferromagnetic Domains and the Giant Planar Hall Effect

The magnetic domains in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  were imaged using a magneto-optical technique that utilizes a Bi-doped garnet layer as an optical magnetic field sensor [29]. The in-plane magnetized garnet layer, with a mirror on one side, is placed directly on the sample, the mirror side adjacent to the sample. Fringing magnetic fields emanating from the sample at domain edges induce Faraday rotation in the garnet, which is visualized in a polarized-light microscope. Figure 4 (left) shows domain images at 15 K taken near a corner of the GaMnAs/GaAs (001) sample for fields applied along the easy [100] direction indicated in frame 4a. In the positively magnetized remanent state (Fig. 4a) the magnetic contrast, seen as white lines, arises only at the sample edges and near a defect at the left edge. With the application of a reversal field (Fig. 4b) a transversely polarized domain nucleates at the top corner of the sample. The moment orientation of this domain is revealed by the contrast at the edges (a more detailed discussion of the orientations will be given below). While the contrast at the right edge remains unchanged, the contrast on the left turns dark, indicating a 90° domain rotation. The domain boundary appears as a jagged bright line, consistent with the indicated moment orientations. The transverse domain expands with increasing field (Fig. 4c). Then a domain totally reversed to the original orientation nucleates at the left edge (Fig. 4d), and spreads rapidly through the entire sample (Figs. 4e and 4f).

Such magnetization reversal through an intermediate (nearly) 90° domain state is typical for samples displaying biaxial magnetic anisotropy. In fact this accounts for the two stages in the magnetization hysteresis for the easy axes seen in the inset in Fig. 2a. Figure 4 (right) shows domain images observed at 35 K. In the remanent states (Figs. 4g and 4l) magnetic contrast appears only at the right edge, indicating a moment orientation along [110]. During the interim process of magnetization reversal, 180° domains nucleate and expand (Figs. 4h–4j), as is typical for magnetic systems with uniaxial in-plane anisotropy. The zig-zag domain boundaries are clearly visible. Contrast at the left [110] edge of the sample appears only in elevated fields (Fig. 4k), when the magnetization is forced into the field direction. Figure 4 thus demonstrates that on increasing the temperature the magnetic anisotropy changes from biaxial (with easy axes approximately along both [100] and [010]) to uniaxial

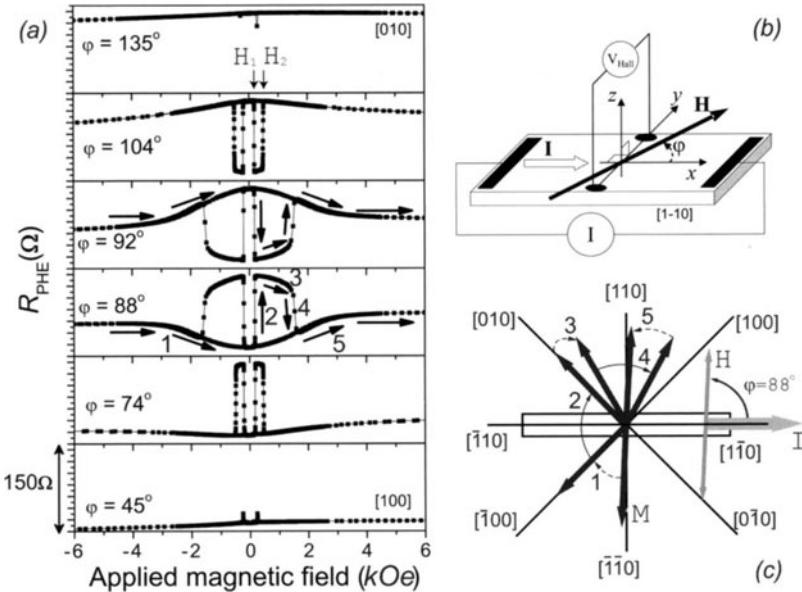


**Fig. 4.** Domain boundaries observed by magneto-optical imaging near a corner of a  $\text{Ga}_{0.97}\text{Mn}_{0.03}\text{As}/\text{GaAs}$  sample at 15 K [left panels: (a) to (f)] and at 35 K [right panels: (g) to (l)] in the indicated fields applied along the [100] direction (vertical axis). The frame width is 1 mm. Fringing field defining a domain boundary with a positive perpendicular component are seen as bright lines; those with negative perpendicular component as dark lines. Black arrows indicate the orientation of the magnetic moments in the domains. The values on the left top corner of each frame indicate the magnitude of the magnetic field at which the frame was observed.

(with easy axis along [110]). Note that, even though the Mn moments are diluted and located randomly, the magnetization reversal proceeds through the nucleation and expansion of large, well-defined domains, several hundred microns on the side.

The magnetization reversal model illustrated above, in combination with anisotropic magnetoresistance data (AMR) [30], can very nicely explain the recently-observed giant planar Hall effect (GPHE) [13,20], which displays large Hall resistance jumps in epitaxial GaMnAs layers when an in-plane magnetic field is swept. Here one should note that planar Hall effect (PHE) is not restricted to GaMnAs films only. It is in fact expected in all ferromagnetic semiconductors in which resistance depends strongly on the relative orientation of magnetization  $\mathbf{M}$  with respect to the direction of the current. Additionally, for GPHE jumps to be observed, the ferromagnetic film needs to have an in-plane easy axis, and the reversal of magnetization must proceed through an intermediate state [13,24].

As an example, in Fig. 5a we show PHE data observed at 1.4 K in a  $\text{Ga}_{0.95}\text{Mn}_{0.05}\text{As}_{0.97}\text{Sb}_{0.03}/\text{GaAs}$  (100) film. The longer edge of the Hall-bar-shaped sample, and hence the direction of the current  $\mathbf{I}$ , was chosen along the  $[1\bar{1}0]$  crystallographic direction (see Fig. 5b). Measurements were performed



**Fig. 5.** (a) Low-field planar Hall effect (PHE) data for various field orientations at 1.8K. (b) A schematic plot of experiment setup for PHE measurements. (c) A schematic magnetization switching process is illustrated for  $\varphi_H = 88^\circ$  as viewed from above.

with the magnetic field  $\mathbf{H}$  confined to the (001) plane, and oriented at various angles  $\varphi$  (from  $45^\circ$  to  $135^\circ$ ) with respect to the current direction  $\mathbf{I}$ . Note that the maximum value of PHE resistance is attained when the magnetization  $\mathbf{M}$  is parallel to the [010] direction (obtained by applying the field  $\mathbf{H}$  along one of the easy axes, at  $\varphi = 135^\circ$ ), and the minimum value is attained for  $\mathbf{M}$  parallel to the [100] direction ( $\mathbf{H}$  along the other easy axis, at  $\varphi = 45^\circ$ ). For all angles  $\varphi$  for which data are presented in Fig. 5a, two-jumps in the PHE resistance are observed as the magnetic field is swept to its opposite value for a given orientation  $\varphi$  (from -6000 Oe to 6000 Oe, or vice versa). Those jumps, at field values  $\pm H_1$  and  $\pm H_2$ , are related to domain nucleation and reversal through the intermediate state. The first switching field,  $H_1$ , is almost independent of the field orientation; and the second switching field,  $H_2$ , attains the largest value when the field is oriented close to the [110] direction ( $\varphi = 90^\circ$ ), but decreases dramatically and approaches the first jump at orientations around  $\varphi = 45^\circ$  ( $\mathbf{H} \parallel [100]$ ).

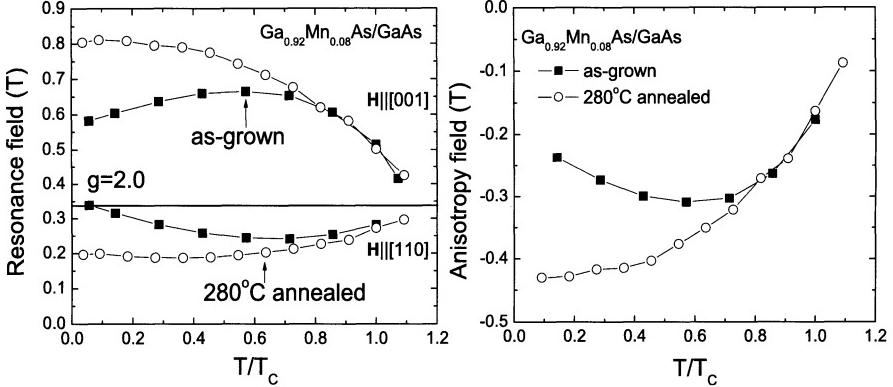
Let us concentrate on the curve for  $88^\circ$  in Fig. 5a and the corresponding schematic plot in Fig. 5c. When the applied field is swept from -6000 Oe to 0 Oe,  $\mathbf{M}$  is rotated from the field direction (which is close to the  $[\bar{1}\bar{1}0]$  direction) toward an easy axis ( $[\bar{1}00]$ ), marked #1 in Fig. 5a and 5c. By increasing the field to a small positive value ( $\sim 190$  Oe), the first resistance jump (marked #2 in Fig. 5a and 5c) is observed at  $H_1$  due to a  $90^\circ$  domain rotation from

[ $\bar{1}00$ ] to the other easy axis, [010]. As the field continues to increase further, the resistance slightly decreases (marked #3) and the direction of  $\mathbf{M}$  tilts to the direction of  $\mathbf{H}$ , eventually reaching the second jump at  $H_2$ ( $\sim$ 1550 Oe), that corresponds to the switching of magnetization from the left side of the field direction to the right side, marked as #4. With further increase of the field,  $\mathbf{M}$  tilts back to the direction of the applied field, oriented close to the [110] direction, marked as #5. These measurements clearly reveal that the MA behavior in the quaternary alloy  $\text{Ga}_{0.95}\text{Mn}_{0.05}\text{As}_{0.97}\text{Sb}_{0.03}$  is dominated by the cubic term at low temperatures, but it is also affected by the small uniaxial anisotropy along [110], which is responsible for the existence of an intermediate state in the magnetization reversal [13,20,22,24].

## 5 Manipulation of Magnetic Anisotropy by Controlling Hole Concentration

It is expected on theoretical grounds that the orientation of the easy-axis can be reoriented in a III-Mn-V system, and can be controlled by varying the temperature, the carrier density  $p$ , or the strain [17]. Sawicki *et al.* [31] have demonstrated by SQUID magnetization measurements that one can induce a spin reorientation from [001] to [100] direction by increasing the temperature in films of GaMnAs grown on GaAs (001) substrate with appropriately low values of  $p$ . Moreover, temperature-induced cross-over of the easy axes have also been found in other alloys, e.g., in AlGaMnAs [32] and InMnAs [33]. In this section, we will review the studies on the manipulation of MA through various post-growth treatments (low temperature annealing) or through structural design (modulation doping). At the end, we will focus on the recently observed reorientation of the easy axis in InMnAs/GaSb heterostructure as a function of either temperature or illumination [8].

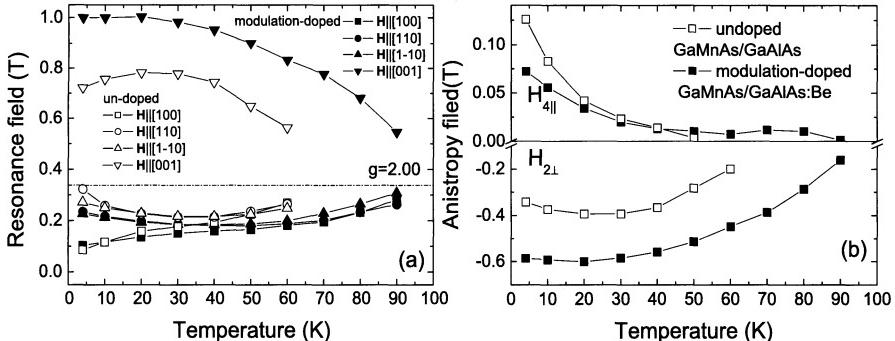
As reported earlier, appropriate annealing can alter (improve) the FM properties of GaMnAs, most notably by increasing its Curie temperature, increasing its magnetic moment, and changing the temperature dependence of its magnetization to a more mean-field-like behavior [34,35]. In Fig. 6 we compare the temperature-dependence of FMR positions for a 120 nm  $\text{Ga}_{0.92}\text{Mn}_{0.08}\text{As}$  film grown on GaAs (100) substrate observed before and after annealing. The annealing was carried out in the atmosphere of  $\text{N}_2$  gas at the temperature of 280°C for 1.0 hour, and subsequently cooled by a rapid quench to room temperature. A remarkable increase of the resonance shifts from  $g = 2.0$  were observed after such heat treatment. The analysis of the data indicates that the observed increase is caused primarily by a dramatic increase in the magnitude of uniaxial anisotropy energy  $K_{2\perp}$ , as seen in the bottom panel of Fig. 6. It is now well established that annealing reduces the concentration of Mn interstitials, thus increasing both the concentration of magnetically-active Mn and of the holes, and eventually improving the homogeneity of the samples. XRD studies have also revealed that the lattice



**Fig. 6.** Top panel: Comparison of the temperature dependences of FMR fields observed at two high-symmetry field orientations in the (1̄10) plane for  $\text{Ga}_{0.92}\text{Mn}_{0.08}\text{As}/\text{GaAs}$  before and after annealing (solid and open symbols, respectively). Bottom panel: Comparison of the anisotropy field  $2K_{2\perp}/M$  before and after annealing for the same sample. The solid lines are guides for the eye.

constant of GaMnAs is reduced after annealing, thus implying a decrease of compressive strain due to the lattice mismatch [36]. The increase of  $K_{2\perp}$  in the presence of the reduction of the build-in compressive strain can only be attributed either to the improvement of the homogeneity, or to an increase of the hole concentration. One should note here that recent SQUID magnetization studies by Sawicki *et al.* [31] performed on GaMnAs samples annealed at various conditions do point to the influence of the hole concentration on magnetic anisotropy.

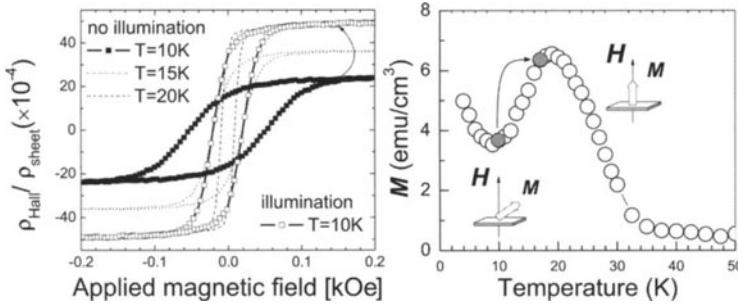
In this context studies of MA in *modulation-doped* GaMnAs layers provide an especially unambiguous and instructive demonstration of the changes of MA with hole concentration. It has recently been found that modulation doping of  $\text{Ga}_{0.76}\text{Al}_{0.24}\text{As}$  barriers with Be acceptors in  $\text{Ga}_{0.94}\text{Mn}_{0.06}\text{As}(5.6 \text{ nm})/\text{GaAlAs}(13.5 \text{ nm})$  heterostructures grown on GaAs (100) substrate (when the doping is carried out *after* GaMnAs has been deposited) leads to an increase of the Curie temperature  $T_C$  of the  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  layer (typically from 70 K in undoped to over 100 K in modulation-doped layers) [37]. The ability to observe in these heterostructures a clear Lorenzian-like FMR peaks up to the Curie temperature has allowed us to use this technique for exploring the dependence of MA on the hole concentration as it is varied by modulation doping [38]. The resonance fields at four specific field orientations, and the anisotropy fields obtained by these experiments ( $H_{2\perp} = 2K_{2\perp}/M$ ,  $H_{4\parallel} = 2K_{4\parallel}/M$ ) at various temperatures are plotted in Fig. 7 for two heterostructures: one undoped (open symbols) and one modulation-doped (solid symbols). As seen in Fig. 7a, remarkable increase of the resonance shifts from  $g = 2.00$  were observed for the modulation-doped sample, indicating a large change of MA. The dependence of the cubic and the uniaxial anisotropy on



**Fig. 7.** Temperature dependence of magnetic properties for an undoped GaMnAs/GaAlAs heterostructure (open symbols) and a modulation-doped GaMnAs/GaAlAs:Be heterostructure (solid symbols). (a) FMR fields observed for four high-symmetry field orientations of the applied magnetic field  $\mathbf{H}$ ; (b) uniaxial magnetic anisotropy fields  $2K_{2\perp}/M$  (bottom panel) and cubic anisotropy fields  $2K_{4\parallel}/M$  (top panel).

the hole concentration are opposite in this concentration range, as seen in Fig. 7b, where the temperature variation of anisotropy fields is presented for both undoped and doped heterostructures. The modulation doping unambiguously increases the value of the perpendicular uniaxial anisotropy  $K_{2\perp}$ , and reduces the in-plane cubic anisotropy  $K_{4\parallel}$ . Note that the compressive strain due to the lattice mismatch between GaMnAs and the GaAs substrate, which influences the anisotropy, is the same in both samples. That is because not only the total concentration of Mn was the same in both specimens, but also the fraction of Mn occupying interstitial positions (and thus having a profound effect on the lattice constant of GaMnAs [36,39]) was not affected by the doping, since Be was introduced into the GaAlAs barrier *after* competing the growth of the GaMnAs layer, and therefore could not lead to the creation of additional Mn interstitials ( $Mn_I$ ).[40] Since  $Mn_I$  also act to reduce the number of ferromagnetically-active Mn spins antiferromagnetically-coupled  $Mn_{Ga}$ - $Mn_I$  pairs (where  $Mn_{Ga}$  stands for substitutional Mn at the Ga site), the fact that the concentration of  $Mn_I$  is unaffected by Be doping also guarantees also guarantees that the concentration of ferromagnetically-active Mn is the same in both samples. Any observed difference between the two specimens must therefore be due to the change in the hole concentration, clearly pointing to the crucial role which holes play in determining the MA of  $Ga_{1-x}Mn_xAs$ . It is important to recall that the theoretical calculations [15,16] indeed do predict that the MA is expected to depend on the hole concentration. A quantitative analysis of this effect is, however, beyond the scope of the present paper.

We finally consider changes in anisotropy in III-Mn-V alloys induced (directly or indirectly) by the temperature – which, as will be argued, may also



**Fig. 8.** Magnetization curves deduced from AHE with and without illumination for an InMnAs/GaSb heterostructure. Magnetization  $M$  as function of temperature  $T$  for the same sample at  $H = 10$  Oe is plotted in the right panel. In all cases, a magnetic field is applied normal to the sample plane.

have their origin in the variation of the hole concentration. For example, temperature-dependence of spin reorientation has recently been observed in FM  $In_{1-x}Mn_xAs/GaSb$  heterostructures [33,8], where  $In_{1-x}Mn_xAs$  is under tensile strain. Additionally, it was demonstrated that light-induced bolometric effects can be used to control the direction of magnetization in the same heterostructure. In spite of the fact that a tensile strain is known to favor an easy axis normal to the layer plane, we have observed that at low temperature the magnetization direction in many  $In_{1-x}Mn_xAs/GaSb$  samples lies close to the plane of the layer, as seen in the right panel of Fig. 8. As the temperature increases, however, the easy axis in this  $In_{1-x}Mn_xAs(10\text{ nm})/GaSb$  heterostructure ( $x = 0.05$ ,  $T_C = 35$  K) grown on GaAs (100) substrate is observed to rotate to the direction normal to the layer plane within a relatively temperature window  $\Delta T \approx 8$  K.

Within this critical temperature range, we demonstrated this by using illumination to tune  $T$  via the bolometric effect. A beam of white light from a standard 100 W bulb was focused onto the InMnAs sample mounted on a cold finger in a closed-cycle helium optical cryostat. The sample temperature was monitored by two thermo-resistors attached to the cold finger above and below the sample, and  $T$ -dependent Hall measurements were carried out with and without illumination. In Fig. 8 we show by solid symbols the magnetization deduced from AHE (assuming  $M \propto \rho_{Hall}/\rho_{sheet}$ , where  $\rho_{Hall}$  and  $\rho_{sheet}$  are the Hall and sheet resistivities, respectively) at 10 K in the dark. The open symbols, representing data observed under illumination, indicate a pronounced *increase* in the amplitude of the AHE (by more than a factor of two), accompanied by a change to a nearly-square hysteresis. The reverse effect is observed when the light is turned off. For comparison, the AHE curves at 15 K (dotted line) and 20 K (dashed line) observed in the dark are also plotted in the figure.

Sawicki *et al.*, who observed temperature-induced anisotropy changes in  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ , suggested that the spin reorientation observed in their experiments reflected the anisotropy of the valence subbands in a zinc-blende semiconductor, which can strongly depend on the hole concentrations, and temperature [31]. It should be emphasized, however, that the  $\text{In}_{1-x}\text{Mn}_x\text{As}/\text{GaSb}$  case is more complicated. The theoretically calculated spin anisotropy [8] implies an easy axis that lies at an angle intermediate between the in-plane [110] direction and the out-of-plane [001] direction ( $41^\circ$  from [110], toward [001]), irrespective of temperature. On the other hand, even a relatively small variation of the hole density can produce a dramatic reorientation of the easy axis [8]. The theoretical simulations just referred to confirm the potential for dramatic reorientations of the magnetization direction, although it remains unclear why the easy axis should be so sensitive to small temperature variations (within a rather narrow window of  $\Delta T \approx 8\text{ K}$ ). One interpretation is that the data reflect a variation of the hole density with  $T$ . although such a variation is unlikely in an isolated InMnAs film, it is possible that the answer lies in the staggered-gap type-II band alignment of the InMnAs/GaSb heterojunction, since in this case some of the holes originating in the InMnAs are transferred to the adjacent GaSb layer due to its higher valence band maximum, thus altering the system by creating an internal electric field together with a density gradient in the InMnAs layer [7,41]. Further study is clearly necessary to understand this interesting and potentially important temperature-dependent process.

## 6 Closing Remarks

In closing, it should be remarked that all the evidence discussed above clearly points to the fact that MA strongly reflects the anisotropy of the hole subbands as shaped by the strain. The geometry and the population of these sub-bands depend on the magnetization (through Zeeman splitting) and on the hole concentration, both of which are functions of temperature, so that the anisotropy is expected to show a dependence on temperature as well. Although the temperature-induced cross-over of the easy axes has been theoretically anticipated from the mean-field Zener model [15,16], the predicted orientations of the easy axes (at least for the case of GaMnAs/GaAs) differ from experimental observations [31]. While the dependence of MA on the hole concentration is complicated and as yet unclear, it is important to appreciate that manipulation of the MA through hole concentration is likely to lead to future device applications. It is therefore especially important to address this issue by further rigorous theoretical studies.

## Acknowledgements

This work was supported by the DARPA/SpinS Program through the Office of Naval Research, by NSF Grant DMR02-10519, and by the U.S. Department of Energy under BES Contract No. W-31-109-ENG-38.

## References

1. H. Ohno, *Science* **281**, 951 (1998).
2. H. Ohno, *J. Magn. Magn. Mater.* **200**, 110 (1999) and reference therein.
3. K. C. Ku, S. J. Potashnik, R. F. Wang, M. J. Seong, E. Johnston-Halperin, R. C. Meyers, S. H. Chun, A. Mascarenhas, A. C. Gossard, D. D. Awschalom, P. Schiffer, and N. Samarth, *Appl. Phys. Lett.* **82**, 2302 (2003).
4. K.W. Edmonds, P. Boguslawski, K.Y. Wang, R.P. Campion, N.R.S. Farley, B.L. Gallagher, C.T. Foxon, M. Sawicki, T. Dietl, M.B. Nardelli, J. Bernholc, *Phys. Rev. Lett.* **92**, 037201 (2004).
5. J. K. Furdyna, X. Liu, Y. Sasaki, S. J. Potashnik, and P. Schiffer, *J. Appl. Phys.* **91**, 7490 (2002).
6. J. K. Furdyna, X. Liu, W. L. Lim, Y. Sasaki, T. Wojtowicz, I. Kuryliszyn, S. Lee, K. M. Yu, and W. Walukiewicz, *J. Korean Physical Society* **42**, S579 (2003).
7. S. Koshihara, A. Oiwa, M. Hirasawa, S. Katsumoto, Y. Iye, C. Urano, H. Takagi, and H. Munekata, *Phys. Rev. Lett.* **78**, 4617 (1997).
8. X. Liu, W. L. Lim, L. V. Titova, T. Wojtowicz, M. Kutrowski, K. J. Yee, M. Dobrowolska, J. K. Furdyna, S. J. Potashnik, M. B. Stone, P. Schiffer, I. Vurgaftman and J. R. Meyer, *Physica E* **20**, 370 (2004).
9. H. Ohno, D. Chiba, F. Matsukura, T. Omiya, E. Abe, T. Dietl, Y. Ohno, and K. Ohtani, *Nature* **408**, 944 (2000).
10. D. Chiba, M. Yamanouchi, F. Matsukura, H. Ohno, *Science* **301**, 943 (2003).
11. G. Schmidt, L. Molenkamp in "Semiconductor Spintronics and Quantum Computing", eds. A. D. Awschalom, D. Loss, N. Samarth (Springer Verlag, Berlin, 2002).
12. D. Chiba, N. Akiba, F. Matsukura, Y. Ohno, and H. Ohno, *Appl. Phys. Lett.* **77**, 1873 (2000); M. Tanaka and Y. Higo, *Phys. Rev. Lett.* **87**, 026602 (2001); S. H. Chun, S. J. Potashnik, K. C. Ku, P. Schiffer, and N. Samarth, *Phys. Rev. B* **66**, 100408 (2002).
13. H. X. Tang, R. K. Kawakami, D. D. Awschalom, and M. L. Roukes, *Phys. Rev. Lett.* **90**, 107201 (2003).
14. C. Rüster, T. Borzenko, C. Gould, G. Schmidt, L. W. Molenkamp, X. Liu, T. J. Wojtowicz, J. K. Furdyna, Z. G. Yu, and M. E. Flatté, *Phys. Rev. Lett.* **91**, 216602 (2003)
15. T. Dietl, H. Ohno, F. Matsukura, J. Cibert, and D. Ferrand, *Science* **287**, 1019 (2000).
16. M. Abolfath, T. Jungwirth, J. Brum, and A.H. MacDonald, *Phys. Rev. B* **63**, 054418 (2001).
17. T. Dietl, H. Ohno, and F. Matsukura, *Phys. Rev. B* **63**, 195205 (2001).
18. A. Shen, A. Oiwa, A. Endo, S. Katsumoto, Y. Iye, H. Ohno, F. Matsukura, Y. Sugawara, N. Akiba and T. Kuroiwa, *J. Cryst. Growth* **175/176**, 1069 (1997).

19. G. P. Moore, J. Ferré, A. Mougin, M. Moreno and L. Däweritz, *J. Appl. Phys.* **94**, 4530 (2003).
20. K. Hamaya, T. Taniyama, Y. Kitamoto, R. Moriya and H. Munekata, *J. Appl. Phys.* **94**, 7657 (2003).
21. X. Liu, Y. Sasaki, J. K. Furdyna, *Phys. Rev. B* **67**, 205204 (2003).
22. U. Welp, V. K. Vlasko-Vlasov, X. Liu, J. K. Furdyna, and T. Wojtowicz, *Phys. Rev. Lett.* **90**, 167206 (2003).
23. S. Katsumoto, A. Oiwa, Y. Iye, H. Ohno, F. Matsukura, A. Shen, and Y. Sugawara, *Phys. Status Solidi (b)* **205**, 115 (1998).
24. D. Hrabovsky, E. Vanelle, A. R. Fert, D. S. Yee, J. P. Redoules, J. Sadowski, J. Kanski and L. Ilver, *Appl. Phys. Lett.* **81**, 2806 (2002).
25. U. Welp, V. K. Vlasko-Vlasov, X. Liu, J. K. Furdyna, and T. Wojtowicz, unpublished.
26. J. J. Krebs, B. T. Jonker, and G. A. Prinz, *J. Appl. Phys.* **61**, 2596 (1987).
27. M. Farle, *Rep. Prog. Phys.* **61**, 755 (1998).
28. P. E. Wigen, *Phys. Rev.* **133**, A1557 (1964).
29. V. K. Vlasko-Vlasov, U. Welp, J. S. Jiang, D. J. Miller, G. W. Crabtree, and S. D. Bader, *Phys. Rev. Lett.* **86**, 4386 (2001); L. A. Dorosinskii, M. V. Indenbom, V. I. Nikitenko, Yu. A. Ossip'yan, A. A. Polyanskii and V. K. Vlasko-Vlasov, *Physica* (Amsterdam) **203C**, 149 (1992).
30. D. V. Baxter, D. Ruzmetov, J. Scherschligt, Y. Sasaki, X. Liu, J. K. Furdyna, and C. H. Mielke, *Phys. Rev. B* **65**, 212407 (2002).
31. M. Sawicki *et al.*, *J. Superconductivity/Novel Magnetism* **16**, 7 (2003); M. Sawicki, *et al.*, to be published in *Phys. Rev. B*.
32. K. Takamura, F. Matsukura, D. Chiba, and H. Ohno, *Appl. Phys. Lett.* **81**, 2590 (2002).
33. T. Endo, T. Slupiński, S. Yanagi, A. Oiwa, and H. Munekata, unpublished.
34. T. Hayashi, Y. Hashimoto, S. Katsumoto, and Y. Iye, *Appl. Phys. Lett.* **78**, 1691 (2001).
35. S. J. Potashnik, K. C. Ku, S. H. Chun, J. J. Berry, N. Samarth, and P. Schiffer, *Appl. Phys. Lett.* **79**, 1495 (2001).
36. I. Kuryliszyn-Kudelska, J. Z. Domagala, T. Wojtowicz, X. Liu, E. Lusakowska, W. Dobrowolski, and J. K. Furdyna, *J. Appl. Phys.* **95**, 603 (2004).
37. T. Wojtowicz, W.L. Lim, X. Liu, M. Dobrowolska, J. K. Furdyna, K. M. Yu, W. Walukiewicz, I. Vurgaftman, and J. R. Meyer, *Appl. Phys. Lett.* **83**, 4220 (2003).
38. X. Liu, W. L. Lim, M. Dobrowolska, J. K. Furdyna, and T. Wojtowicz, unpublished.
39. J. Mašek, J. Kudrnovský, and F. Máca, *Phys. Rev. B* **67**, 153203 (2003).
40. K. M. Yu, W. Walukiewicz, T. Wojtowicz, W.L. Lim, X. Liu, M. Dobrowolska, and J. K. Furdyna, unpublished.
41. A. Oiwa, T. Slupiński, and H. Munekata, *Appl. Phys. Lett.* **78**, 518 (2001).

## Part IX

### Contact and Interface Quantum Phenomena

# Singlet-Triplet Mixing in Superconductor-Ferromagnet Hybrid Devices

Matthias Eschrig<sup>1</sup>, J. Kopu<sup>1,2</sup>, A. Konstandin<sup>1</sup>,  
J. C. Cuevas<sup>1</sup>, M. Fogelström<sup>3</sup>, and Gerd Schön<sup>1,4</sup>

<sup>1</sup> Institut für Theoretische Festkörperphysik, Universität Karlsruhe,  
76128 Karlsruhe, Germany

<sup>2</sup> Low Temperature Laboratory, Helsinki University of Technology,  
FIN-02015 HUT, Finland

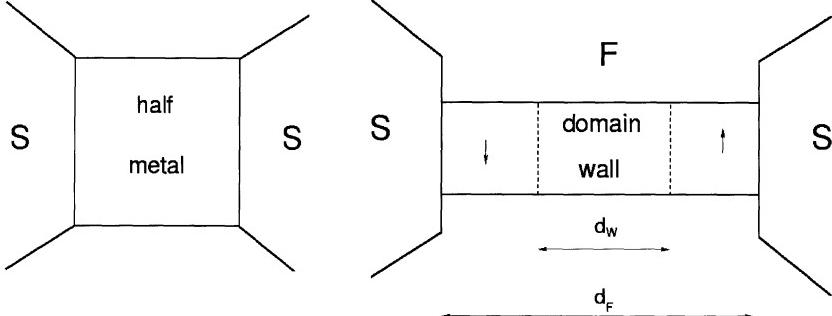
<sup>3</sup> Applied Quantum Physics, MC2, Chalmers,  
S-41296 Göteborg, Sweden

<sup>4</sup> Institut für Nanotechnologie, Forschungszentrum Karlsruhe,  
76021 Karlsruhe, Germany

**Abstract.** We develop a theory which describes hybrid structures consisting out of superconducting and ferromagnetic parts. We give two examples for applications. First, we consider a hybrid structure containing a strong ferromagnet in the ballistic limit. Second, we study for a weak ferromagnet the influence of a domain wall on the superconducting proximity effect. In both cases we account quantitatively for the mixing between singlet and triplet correlations.

## 1 Introduction

Hybrid structures containing ferromagnetic materials became recently a focus of nano electronic research because of their relevance for spintronics applications. Consequently, it is desirable to understand how in the case of a superconductor coupled to a ferromagnetic material superconducting correlations penetrate into the ferromagnet. A powerful method to treat such problems is the quasi-classical theory of superconductivity developed by Larkin and Ovchinnikov and by Eilenberger [1,2]. Within this theory the quasiparticle motion is treated on a classical level, whereas the particle-hole and the spin degree of freedoms are treated quantum mechanically. The paper consist of two parts. First we demonstrate our method for a ballistic heterostructure containing a *strong* ferromagnet, for which we chose for simplification a completely polarized ferromagnet, a half metal. We show results for the modification of the quasiparticle density of states due to the proximity effect in the half-metallic material in a superconductor/half-metal/superconductor device (see Fig. 1, left). Second, we study for a diffusive heterostructure containing a *weak* ferromagnet the influence of a domain wall of the Bloch type on the quasiparticle spectrum. The corresponding setup is shown on the right in Fig. 1. We show results for different ratios between the domain wall width and the superconducting coherence length.



**Fig. 1.** Schematic picture of the studied devices. On the left, a ballistic heterostructure where a half metal is sandwiched between two superconductors. Here, the conventional proximity effect is completely suppressed as a result of the complete spin polarization of the half metal. On the right, a diffusive weak ferromagnet between two superconductors. The ferromagnet contains a Bloch domain wall of size  $d_w$

## 2 Basic Equations: Ballistic Case

### 2.1 Transport Equations

The central quantity in quasi-classical theory of superconductivity [1,2] is the quasi-classical Green function  $\check{g}(\mathbf{p}_F, \mathbf{R}, E, t)$  that depends on the spatial coordinate  $\mathbf{R}$  and time  $t$ . It describes quasiparticles with energy  $E$  (measured from the chemical potential) and Fermi momentum  $\mathbf{p}_F$  moving along classical trajectories with direction given by the Fermi velocity  $\mathbf{v}_F(\mathbf{p}_F)$ .[3] The quasi-classical Green function is a functional of self energies  $\check{\Sigma}(\mathbf{p}_F, \mathbf{R}, E, t)$ , which in general include molecular fields, the superconducting order parameter  $\Delta(\mathbf{p}_F, \mathbf{R}, t)$ , impurity scattering, and external fields. The quantum mechanical degrees of freedom of the quasiparticles show up in the matrix structure of the quasi-classical propagator and the self energies. It is convenient to formulate the theory using  $2 \times 2$  matrices in Keldysh space (denoted by a “check” accent), the elements of which in turn are  $4 \times 4$  Nambu-Gor’kov matrices in combined particle-hole (denoted by a “hat” accent) and spin space. The structure of the propagators and self energies in Keldysh-space and particle-hole space is as follows,

$$\check{g} = \begin{pmatrix} \hat{g}^R & \hat{g}^K \\ 0 & \hat{g}^A \end{pmatrix}, \quad \hat{g}^{R,A} = \begin{pmatrix} g^{R,A} & f^{R,A} \\ \tilde{f}^{R,A} & \tilde{g}^{R,A} \end{pmatrix}, \quad \hat{g}^K = \begin{pmatrix} g^K & f^K \\ -\tilde{f}^K & -\tilde{g}^K \end{pmatrix}, \quad (1)$$

$$\check{\Sigma} = \begin{pmatrix} \hat{\Sigma}^R & \hat{\Sigma}^K \\ 0 & \hat{\Sigma}^A \end{pmatrix}, \quad \hat{\Sigma}^{R,A} = \begin{pmatrix} \Sigma^{R,A} & \Delta^{R,A} \\ \tilde{\Delta}^{R,A} & \tilde{\Sigma}^{R,A} \end{pmatrix}, \quad \hat{\Sigma}^K = \begin{pmatrix} \Sigma^K & \Delta^K \\ -\tilde{\Delta}^K & -\tilde{\Sigma}^K \end{pmatrix}. \quad (2)$$

The elements of the  $2 \times 2$  Nambu-Gor’kov matrices are  $2 \times 2$  matrices in spin space, e.g.  $g^R = g_{\alpha\beta}^R$  with  $\alpha, \beta = \{\uparrow, \downarrow\}$ , and similarly for others. In

writing Eqs. (1) and (2) we used general symmetries, which are accounted for by the “tilde” operation,

$$\tilde{X}(\mathbf{p}_F, \mathbf{R}, E, t) = X(-\mathbf{p}_F, \mathbf{R}, -E, t)^*. \quad (3)$$

The quasi-classical Green functions satisfy the Eilenberger-Larkin-Ovchinnikov transport equation and normalization condition

$$[E\check{\tau}_3 - \check{\Sigma}, \check{g}]_{\otimes} + i\mathbf{v}_F \cdot \nabla \check{g} = 0, \quad \check{g} \otimes \check{g} = -\pi^2 \check{1}. \quad (4)$$

The non-commutative product  $\otimes$  combines matrix multiplication with a convolution over the internal variables, and  $\check{\tau}_3 = \hat{\tau}_3 \check{1}$  is a Pauli matrix in particle-hole space.

The functional dependence of the quasi-classical propagator on the self energies is given in the form of self-consistency conditions. For instance, for a weak-coupling, *s*-wave order parameter the condition reads

$$\hat{\Delta}(\mathbf{R}, t) = \lambda \int_{-E_c}^{E_c} \frac{dE}{4\pi i} \langle \hat{f}^K(\mathbf{p}_F, \mathbf{R}, E, t) \rangle_{\mathbf{p}_F}, \quad (5)$$

where  $\lambda$  is the strength of the pairing interaction, and  $\langle \rangle_{\mathbf{p}_F}$  denotes averaging over the Fermi surface. The cut-off energy  $E_c$  is to be eliminated in favor of the transition temperature in the usual manner.

When the quasi-classical Green function has been determined, physical quantities of interest can be calculated. For instance, the equilibrium local density of states at position  $\mathbf{R}$  for quasiparticles with momentum  $\mathbf{p}_F$  and energy  $E$  (measured from the Fermi level) reads

$$N(\mathbf{p}_F, \mathbf{R}, E) = N_F \frac{1}{-4\pi i} \text{Tr} [\hat{\tau}_3 \hat{g}^R(\mathbf{p}_F, \mathbf{R}, E) - \hat{\tau}_3 \hat{g}^A(\mathbf{p}_F, \mathbf{R}, E)]. \quad (6)$$

where  $N_F$  is the density of states on the Fermi surface.

For heterostructures, the above equations must still be supplemented with boundary conditions at the interfaces. As these conditions are non-trivial in quasi-classical theory, we present these conditions in the following chapter in detail.

## 2.2 Boundary Conditions

In order to formulate the boundary conditions at an interface between two materials, we define first an auxiliary propagator on each side of the interface [4]. Then we relate this auxiliary propagator to the full propagator via a transfer matrix. This approach is completely equivalent to a full scattering matrix approach as we will show at the end of this paragraph. However, the repeated Andreev scattering processes across the interface complicate the full scattering matrix approach, and we found the current method much easier to implement and numerically more stable [5,6].

We use for the auxiliary propagator the notation  $g_o^{\alpha,0}$  and  $\check{g}_i^{\alpha,0}$ , where the upper index  $\alpha = \{l, r\}$  determines the side of the interface (left or right). The lower index denotes the direction of the Fermi velocity. *Incoming* momenta (index  $i$ ) are those with a Fermi velocity pointing towards the interface, and *outgoing* momenta (index  $o$ ) are those with a Fermi velocity pointing away from the interface. We formulate here the boundary conditions for clean surfaces, which conserve the parallel component of the Fermi momentum,  $\mathbf{p}_{\parallel}$ . The auxiliary propagators as function of  $\mathbf{p}_{\parallel}$ , energy  $E$ , and time  $t$  are solutions of the quasi-classical transport equation with the *exact* self energies, together with the normalization condition, however subject to the *auxiliary boundary conditions*,

$$\check{g}_o^{\alpha,0}(\mathbf{p}_{\parallel}, E, t) = \hat{S}_{oi}^{\alpha}(\mathbf{p}_{\parallel}) \check{g}_i^{\alpha,0}(\mathbf{p}_{\parallel}, E, t) \hat{S}_{io}^{\alpha}(\mathbf{p}_{\parallel}), \quad (7)$$

where the following symmetries hold,

$$\hat{S}_{io}^{\alpha}(\mathbf{p}_{\parallel}) = \hat{S}_{oi}^{\alpha}(\mathbf{p}_{\parallel})^{\dagger} = \hat{S}_{oi}^{\alpha}(\mathbf{p}_{\parallel})^{-1}. \quad (8)$$

These boundary conditions are formally equivalent to boundary conditions for an impenetrable interface, however with a surface scattering matrix  $\hat{S}_{oi}^{\alpha}$  determined from the reflection amplitudes of the *full* scattering matrix as explained below. Once these auxiliary propagators are obtained, the full propagators can be obtained directly, without further solving the transport equation, in the following way. We define hopping amplitudes  $\hat{\tau}_{io}^{lr}$ ,  $\hat{\tau}_{io}^{rl}$ ,  $\hat{\tau}_{oi}^{lr}$  and  $\hat{\tau}_{oi}^{rl}$ , for the four channels (left incoming to right outgoing, right incoming to left outgoing, left outgoing to right incoming, and right outgoing to left incoming) which conserve the momentum component parallel to the interface,  $\mathbf{p}_{\parallel}$ . With the help of these amplitudes we solve for the *transfer matrices*,  $\check{t}_i^{\alpha}$ , for incoming trajectories from the following equations,

$$\begin{aligned} \check{t}_i^{\alpha}(\mathbf{p}_{\parallel}, E, t) &= \hat{\tau}_{io}^{\alpha\beta}(\mathbf{p}_{\parallel}) \check{g}_o^{\beta,0}(\mathbf{p}_{\parallel}, E, t) \hat{\tau}_{oi}^{\beta\alpha}(\mathbf{p}_{\parallel}) \\ &\otimes \left( \check{1} + \check{g}_i^{\alpha,0}(\mathbf{p}_{\parallel}, E, t) \otimes \check{t}_i^{\alpha}(\mathbf{p}_{\parallel}, E, t) \right), \end{aligned} \quad (9)$$

where  $(\alpha\beta) = \{(l, r), (r, l)\}$ . The corresponding transfer matrices for outgoing trajectories are related to the ones for incoming trajectories through the relations

$$\check{t}_o^{\alpha}(\mathbf{p}_{\parallel}, E, t) = \hat{S}_{oi}^{\alpha}(\mathbf{p}_{\parallel}) \check{t}_i^{\alpha}(\mathbf{p}_{\parallel}, E, t) \hat{S}_{io}^{\alpha}(\mathbf{p}_{\parallel}). \quad (10)$$

Particle conservation requires certain symmetries between the hopping elements, which are,

$$\hat{\tau}_{oi}^{\beta\alpha}(\mathbf{p}_{\parallel}) = \hat{\tau}_{io}^{\alpha\beta}(\mathbf{p}_{\parallel})^{\dagger} = \hat{S}_{oi}^{\beta}(\mathbf{p}_{\parallel}) \hat{\tau}_{io}^{\beta\alpha}(\mathbf{p}_{\parallel}) \hat{S}_{oi}^{\alpha}(\mathbf{p}_{\parallel}). \quad (11)$$

Consequently, only one amplitude, e.g.  $\hat{\tau}_{io}^{lr}$ , contains free material parameters, the other three amplitudes depend on it. The hopping amplitudes are defined via the transmission amplitudes of the full scattering matrix as is shown

below. Formally, they describe the modifications of the decoupled problem due to virtual hopping processes to the opposite side.

The *full propagators*, fulfilling the desired boundary conditions at the interface, can now be easily calculated. For incoming trajectories they are obtained from

$$\begin{aligned} \check{g}_i^\alpha(\mathbf{p}_\parallel, E, t) &= \check{g}_i^{\alpha,0}(\mathbf{p}_\parallel, E, t) \\ &+ \left( \check{g}_i^{\alpha,0}(\mathbf{p}_\parallel, E, t) + i\pi\check{1} \right) \otimes \check{t}_i^\alpha(\mathbf{p}_\parallel, E, t) \otimes \left( \check{g}_i^{\alpha,0}(\mathbf{p}_\parallel, E, t) - i\pi\check{1} \right), \end{aligned} \quad (12)$$

and for outgoing trajectories from

$$\begin{aligned} \check{g}_o^\alpha(\mathbf{p}_\parallel, E, t) &= \check{g}_o^{\alpha,0}(\mathbf{p}_\parallel, E, t) \\ &+ \left( \check{g}_o^{\alpha,0}(\mathbf{p}_\parallel, E, t) - i\pi\check{1} \right) \otimes \check{t}_o^\alpha(\mathbf{p}_\parallel, E, t) \otimes \left( \check{g}_o^{\alpha,0}(\mathbf{p}_\parallel, E, t) + i\pi\check{1} \right). \end{aligned} \quad (13)$$

In this formulation, the boundary problem effectively reduces to calculating the auxiliary Green functions for perfectly reflecting interfaces. Numerically this is an extremely simple task, *e.g.*, employing the procedure of Riccati parameterization as explained below. Afterwards the boundary Green functions for the partially transmitting interface can be obtained directly from Eqs. (12) and (13), since solving for the necessary transfer matrices (9) only involves a matrix inversion.

In the transfer-matrix description, the phenomenological parameters containing the microscopic information of the interface are the two surface scattering matrices  $S_{oi}^{l,r}$  and the hopping amplitude  $\tau_{io}^{lr}$ . All three quantities are  $2 \times 2$  matrices in spin space. As  $S_{io}^\alpha$  is unitary, it depends (apart from the direction of the quantitation axis) on two parameters, a scalar scattering phase and a spin mixing angle (or, equivalently, a spin rotation angle). Similarly, as  $\tau_{io}^{lr}$  is hermitian, it also depends on two parameters, one describing spin conserving transmission and the other spin flip transmission. All remaining quantities defined above are related to these material parameters by symmetries.

The particle-hole structures of the surface scattering matrix and the hopping amplitude are given by,

$$\hat{S}_{oi}^\alpha(\mathbf{p}_\parallel) = \begin{pmatrix} S_{oi}^\alpha(\mathbf{p}_\parallel) & 0 \\ 0 & \tilde{S}_{oi}^\alpha(\mathbf{p}_\parallel) \end{pmatrix}, \quad \hat{\tau}_{io}^{\alpha\beta}(\mathbf{p}_\parallel) = \begin{pmatrix} \tau_{io}^{\alpha\beta}(\mathbf{p}_\parallel) & 0 \\ 0 & \tilde{\tau}_{io}^{\alpha\beta}(\mathbf{p}_\parallel) \end{pmatrix}, \quad (14)$$

with the hole components,

$$\tilde{S}_{oi}^\alpha(\mathbf{p}_\parallel) = S_{oi}^\alpha(-\mathbf{p}_\parallel)^* = S_{oi}^\alpha(-\mathbf{p}_\parallel)^{tr} \quad (15)$$

$$\tilde{\tau}_{io}^{\alpha\beta}(\mathbf{p}_\parallel) = \tau_{oi}^{\alpha\beta}(-\mathbf{p}_\parallel)^* = S_{oi}^\alpha(-\mathbf{p}_\parallel)^* \tau_{io}^{\alpha\beta}(-\mathbf{p}_\parallel)^* S_{oi}^\beta(-\mathbf{p}_\parallel)^*. \quad (16)$$

Finally, we relate the parameters  $S_{oi}^\alpha$  and  $\tau_{io}^{\alpha\beta}$  to the *full normal state scattering matrix*  $\hat{\mathbf{S}}$ ,

$$\hat{\mathbf{S}} = \begin{pmatrix} \hat{\mathbf{S}}_{ll} & \hat{\mathbf{S}}_{lr} \\ \hat{\mathbf{S}}_{rl} & -\hat{\mathbf{S}}_{rr} \end{pmatrix}. \quad (17)$$

The scattering matrix is diagonal in particle-hole space, with diagonal components

$$\hat{\mathbf{S}}_{\alpha\alpha} = (1 + \pi^2 \hat{\tau}_{oi}^{\alpha\beta} \hat{\tau}_{io}^{\beta\alpha})^{-1} (1 - \pi^2 \hat{\tau}_{oi}^{\alpha\beta} \hat{\tau}_{io}^{\beta\alpha}) \hat{S}_{oi}^{\alpha}, \quad (18)$$

and off-diagonal components

$$\hat{\mathbf{S}}_{\alpha\beta} = (1 + \pi^2 \hat{\tau}_{oi}^{\alpha\beta} \hat{\tau}_{io}^{\beta\alpha})^{-1} 2\pi \hat{\tau}_{oi}^{\alpha\beta}. \quad (19)$$

These identities serve as a precise definition of the auxiliary parameters of the theory,  $S_{oi}^l(\mathbf{p}_\parallel)$ ,  $S_{oi}^r(\mathbf{p}_\parallel)$ , and  $\tau_{io}^{lr}(\mathbf{p}_\parallel)$  in terms of the physical parameters of the full scattering matrix.

## 2.3 Riccati Parameterization

The method of the Riccati parameterization [7,8,9] of the quasi-classical Green functions has proved powerful during recent years. It accounts automatically for the normalization condition. A corresponding parameterization for distribution functions was applied to the Keldysh part of the Green functions. The combined equations in their general form are [9]

$$\begin{aligned} \hat{g}^K &= -2i\pi \hat{N}^R \otimes \begin{pmatrix} (x - \gamma^R \otimes \tilde{x} \otimes \tilde{\gamma}^A) & -(\gamma^R \otimes \tilde{x} - x \otimes \gamma^A) \\ -(\tilde{\gamma}^R \otimes x - \tilde{x} \otimes \tilde{\gamma}^A) & (\tilde{x} - \tilde{\gamma}^R \otimes x \otimes \gamma^A) \end{pmatrix} \otimes \hat{N}^A, \\ \hat{g}^{R,A} &= \pm i\pi \hat{N}^{R,A} \otimes \begin{pmatrix} (1 + \gamma^{R,A} \otimes \tilde{\gamma}^{R,A}) & 2\gamma^{R,A} \\ -2\tilde{\gamma}^{R,A} & -(1 + \tilde{\gamma}^{R,A} \otimes \gamma^{R,A}) \end{pmatrix}, \end{aligned} \quad (20)$$

with

$$\hat{N}^{R,A} = \begin{pmatrix} (1 - \gamma^{R,A} \otimes \tilde{\gamma}^{R,A})^{-1} & 0 \\ 0 & (1 - \tilde{\gamma}^{R,A} \otimes \gamma^{R,A})^{-1} \end{pmatrix}. \quad (21)$$

Thus, the problem is reduced to the solution for two  $2 \times 2$  matrices in spin space,  $\gamma^R$  and  $x$ . Using fundamental symmetries (particle-hole, retarded-advanced) like Eq. (3) and

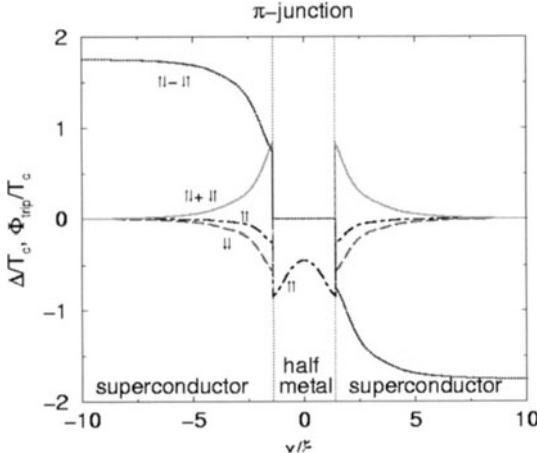
$$\gamma^A(\mathbf{p}_F, \mathbf{R}, E, t) = \gamma^R(-\mathbf{p}_F, \mathbf{R}, -E, t)^{tr}, \quad (22)$$

where  $tr$  denotes the spin-matrix transpose operation, the full retarded, advanced and Keldysh Green functions are obtained from  $\gamma^R$  and  $x$ . The transport equations for the functions  $\gamma^R(\mathbf{p}_F, \mathbf{R}, E, t)$  and  $x(\mathbf{p}_F, \mathbf{R}, E, t)$  are

$$2E\gamma^R + i\mathbf{v}_F \nabla \gamma^R = \gamma^R \otimes \tilde{\Delta}^R \otimes \gamma^R + \Sigma^R \otimes \gamma^R - \gamma^R \otimes \tilde{\Sigma}^R - \Delta^R, \quad (23)$$

and

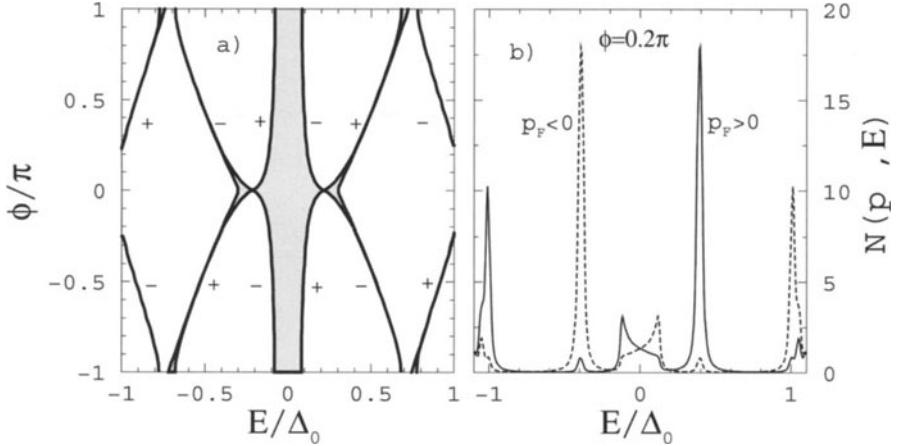
$$\begin{aligned} i\partial_t x + i\mathbf{v}_F \nabla x + (-\gamma^R \otimes \tilde{\Delta}^R - \Sigma^R) \otimes x + x \otimes (-\Delta^A \otimes \tilde{\gamma}^A + \Sigma^A) \\ = -\gamma^R \otimes \tilde{\Sigma}^K \otimes \tilde{\gamma}^A + \Delta^K \otimes \tilde{\gamma}^A + \gamma^R \otimes \tilde{\Delta}^K - \Sigma^K. \end{aligned} \quad (24)$$



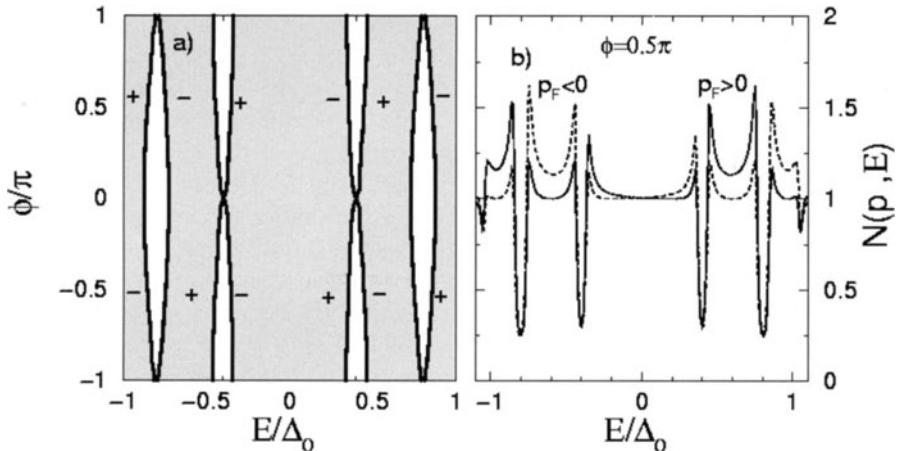
**Fig. 2.** Self consistent order parameter and triplet correlations in an S/HM/S  $\pi$ -junction. The calculations are for temperature  $T = 0.05T_c$ , and for  $\tau_{\downarrow,\uparrow}/\tau_{\uparrow,\uparrow} = 0.7$

### 3 Results: Ballistic Case

As an example we consider here the Andreev quasiparticle spectrum in a half-metallic ferromagnet between two singlet superconductors. This corresponds to the device shown on the left side in Fig. 1. In this case, on the half-metallic side only quasiparticles with one spin direction with respect to the quantization axis are itinerant. The stable equilibrium configuration of such a system is a  $\pi$ -junction, in which the phase of the singlet order parameter on both sides of the interface differs by  $\pi$  [5]. Characteristic of such a heterostructure is the presence of triplet correlations near the interface. Shown in Fig. 2 are the singlet order parameter and all three triplet components which are induced near the interface by the spin rotation effect [10]. The spin rotation effect is taken into account via a surface scattering matrix  $\hat{S}_{oi} = \exp(i\theta\sigma_z/2)\hat{1}$  at the superconducting side of the interface, where  $\theta$  defines a spin-rotation angle and  $\sigma_z$  denotes the Pauli spin matrix [10,11]. Generally, the value of  $\theta$  depends on the angle of impact,  $\psi$  [10] and can approach values of the order of  $\pi$  for strong band splitting [12]. For definiteness, we present results for  $\theta = 0.75\pi \cos\psi$ . On the half-metallic side the scattering matrix has no spin structure. Further, we use the hopping amplitude  $\tau_{io} = (1 + S_{io})\tau_0 \cos\psi$ , where  $\tau_0 = (\tau_{\uparrow,\uparrow}, \tau_{\downarrow,\uparrow})^T$  is determined by the two spin scattering channels from the superconductor to the half-metallic spin-up band. This reflects a spin rotation during transmission which is half of the spin rotation during reflection. The  $\cos\psi$  factor accounts for the reduced transmission at large impact angles. We present results for  $\tau_{\downarrow,\uparrow}/\tau_{\uparrow,\uparrow} = 0.7$  and 0.1,  $2\pi\tau_{\uparrow,\uparrow} = 1.0$ . The length of the half metallic region is  $3\xi_0$  with the coherence length  $\xi_0 = v_f/2\pi T_c$ . We assume cylindrical Fermi surfaces.



**Fig. 3.** Density of states at  $T = 0.05T_c$  for quasiparticles with normal impact at the half-metallic side of the left interface, for  $\tau_{\downarrow,\uparrow}/\tau_{\uparrow,\uparrow} = 0.7$ . a) Dispersion of the maxima of the density of states as function of phase difference. Gapped regions are white, ungapped regions gray. The signs indicate the direction of the current carried by the Andreev states. b) As an example we show spectra for a selected phase difference



**Fig. 4.** The same as Fig. 3 for  $\tau_{\downarrow,\uparrow}/\tau_{\uparrow,\uparrow} = 0.1$

We have studied the behavior of such a heterostructure as a Josephson device. If a phase difference  $\phi$  in the order parameter is present between the left and right end of the heterostructure, the quasiparticle spectrum in the half metal is modified. In Figs. 3 and 4 we show results of our calculations for two sets of parameters.

The spectra consist out of Andreev quasiparticle bands, separated by gaps. In Figs. 3 a) and 4 a) we show these gapped-regions (white) and the non-gapped regions (grey) of the density of states at the half-metallic side of the left interface as a function of the phase difference. One characteristic spectrum for fixed phase difference for both positive and negative momentum direction is shown in Figs. 3 b) and 4 b). The dispersion of the maxima in the density of states with phase difference determines the sign of the Josephson current through the device. We indicate in Figs. 3a) and 4a) by + and – Andreev states, which carry current in positive and negative direction respectively. It can be seen that the sign of the Josephson current corresponds to the sign of  $dE(\phi)/d\phi$ , where  $E(\phi)$  denotes the dispersion of the maxima in the density of states. This is analogous to the current carried by a bound state  $E_b$  dispersing with phase  $\phi$ , which is given by  $J_b \sim (dE_b/d\phi)n(E_b)$ , where  $n(E_b)$  is the equilibrium fermion distribution function.

## 4 Basic Equations: Diffusive Case

### 4.1 Transport Equations

The fundamental quantity for diffusive transport is the Usadel Green function [13], which is the momentum average of the quasi-classical Green function  $\check{g}(\mathbf{R}, E, t) = \langle \check{g}(\mathbf{p}_F, \mathbf{R}, E, t) \rangle_{\mathbf{p}_F}$ . It is a functional of momentum averaged self energies  $\check{\Sigma}(\mathbf{R}, E, t) = \langle \check{\Sigma}(\mathbf{p}_F, \mathbf{R}, E, t) \rangle_{\mathbf{p}_F}$ . The structures of  $\check{g}$  and  $\check{\Sigma}$  are the same as in Eqs. (1) and (2). Eq. (3) is replaced by

$$\tilde{X}(\mathbf{R}, E, t) = X(\mathbf{R}, -E, t)^*. \quad (25)$$

The Usadel Green function obeys the following transport equation and normalization condition [13],

$$[E \hat{\tau}_3 \check{1} - \check{\Sigma}, \check{g}]_{\otimes} + \frac{D}{\pi} \nabla_j (\check{g} \otimes \nabla_j \check{g}) = 0, \quad \check{g} \otimes \check{g} = -\pi^2 \check{1}. \quad (26)$$

Here, summation over the repeated index  $j$  is implied.  $\check{\Sigma}$  does not contain the non-magnetic impurity scattering self energy anymore. We will solve these equations by using the Riccati parameterization technique and present below the resulting equations for the diffusive case. The corresponding equations in a finite vector potential are obtained by applying a gauge transformation to Eqs. (26) (and to Eqs. (28), (29) below).

### 4.2 Boundary Conditions

In the diffusive limit the boundary conditions are formulated in terms of the momentum averaged transfer matrices and Green functions. The resulting

boundary conditions for spin-conserving interfaces were found by Nazarov [14] and follow from the transfer-matrix approach outlined above [6]. They are

$$\begin{aligned}\sigma_\beta \check{g}^\beta \frac{d}{dz} \check{g}^\beta &= \frac{1}{SR_b} \frac{2\pi^2 T[\check{g}^\beta, \check{g}^\alpha]}{4\pi^2 - T(\{\check{g}^\beta, \check{g}^\alpha\} + 2\pi^2)} \\ \sigma_\beta \check{g}^\beta \frac{d}{dz} \check{g}^\beta &= \sigma_\alpha \check{g}^\alpha \frac{d}{dz} \check{g}^\alpha,\end{aligned}\quad (27)$$

where  $z$  is the coordinate along the interface normal,  $\sigma_{\alpha(\beta)}$  and  $\check{g}^{\alpha(\beta)}$  refers to the conductivity and the Keldysh Green function on side  $\alpha(\beta)$  of the interface,  $S$  is the surface area of the contact,  $R_b$  the boundary resistance and  $T$  the transmission probability,  $T = \frac{4\pi^2|\tau|^2}{(1+\pi^2|\tau|^2)^2}$ .

### 4.3 Riccati Parameterization

We use the same parameterization, Eqs. (20)–(21), for the momentum averaged Green's functions. Note that the such defined diffusive quantities  $\gamma^R(\mathbf{R}, E, t)$  and  $x(\mathbf{R}, E, t)$  are by no means in a simple way related to the ballistic quantities  $\gamma^R(\mathbf{p}_F, \mathbf{R}, E, t)$  and  $x(\mathbf{p}_F, \mathbf{R}, E, t)$ . Consequently, also the structure of the transport equations is very different. In particular the gradient terms in the transport equations for these quantities are modified with respect to the gradient terms in the transport equations for the ballistic case. Instead of Eqs. (23) and (24) we obtain for the diffusive case

$$\begin{aligned}2E\gamma^R - iD\left[\nabla^2\gamma^R + (\nabla_j\gamma^R) \otimes \frac{\tilde{f}^R}{i\pi} \otimes (\nabla_j\gamma^R)\right] \\ = \gamma^R \otimes \tilde{\Delta}^R \otimes \gamma^R + \Sigma^R \otimes \gamma^R - \gamma^R \otimes \tilde{\Sigma}^R - \Delta^R,\end{aligned}\quad (28)$$

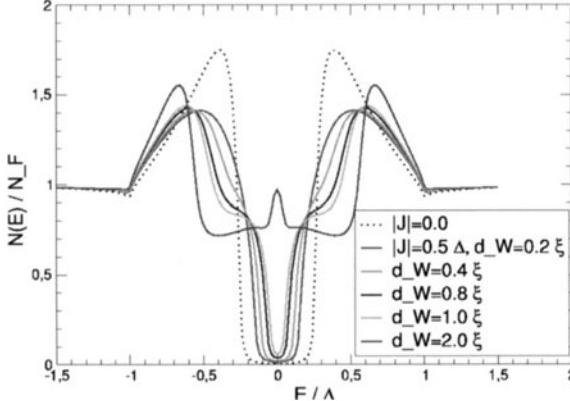
and

$$\begin{aligned}i\partial_t x - iD\left[\nabla^2 x - (\nabla_j\gamma^R) \otimes \frac{\tilde{g}^K}{i\pi} \otimes (\nabla_j\tilde{\gamma}^A)\right. \\ \left. + (\nabla_j\gamma^R) \otimes \frac{\tilde{f}^R}{i\pi} \otimes (\nabla_jx) + (\nabla_jx) \otimes \frac{f^A}{i\pi} \otimes (\nabla_j\tilde{\gamma}^A)\right] \\ + (-\gamma^R \otimes \tilde{\Delta}^R - \Sigma^R) \otimes x + x \otimes (-\Delta^A \otimes \tilde{\gamma}^A + \Sigma^A) \\ = -\gamma^R \otimes \tilde{\Sigma}^K \otimes \tilde{\gamma}^A + \Delta^K \otimes \tilde{\gamma}^A + \gamma^R \otimes \tilde{\Delta}^K - \Sigma^K.\end{aligned}\quad (29)$$

Summation over the repeated index  $j$  is implied. The expressions for  $\tilde{f}^R$ ,  $f^A$ , and  $\tilde{g}^K$  are obtained by comparing Eq. (1) with Eqs. (20)–(21).

## 5 Results: Diffusive Case

As an example we study for this case the heterostructure shown on the right side of Fig. 1, which contains a weak diffusive ferromagnet between two superconductors. A domain wall of Bloch type is centered in the ferromagnetic



**Fig. 5.** Local density of states as a function of energy in the center of a heterostructure containing a weak ferromagnet ( $J = 0.5\Delta$ ) of length  $d_F = 2\xi_0$  between two superconductors. A Bloch domain wall of length  $d_W$  is centered in the ferromagnet (see Fig. 1). Results for various  $d_W$  are shown

part. Problems involving domain walls have previously been treated only with techniques that do not allow access to information on spatial variations [15], or elaborate methods involving rotating coordinate systems [16]. However, since the Riccati description contains the full  $2 \times 2$  spin structure, it extends easily to account for any such situation.

We adopt for the description of a weak ferromagnetic materials ( $J$  much smaller than the Fermi energy) a quasiparticle dispersion given by a spin-dependent energy shift  $E\hat{\tau}_3 \rightarrow E\hat{\tau}_3 - \mathbf{J}(\mathbf{R}) \cdot \boldsymbol{\sigma}$ , in Eq. (26). Here  $\mathbf{J}(\mathbf{R})$  denotes the effective exchange field of the ferromagnet, and  $\boldsymbol{\sigma}$  denotes the vector of Pauli spin matrices. We solve the Usadel equations using the Riccati parameterization for a ferromagnetic material, which contains a domain wall of the Bloch type. We model the domain wall by a  $\mathbf{J}$ -vector rotating across the domain wall with the polar angle  $\theta_J = \text{atan}(z/d_W)$ , where  $d_W$  parameterizes the width of the domain wall.

In Fig. 5 we present results for the local density of states at the center position of the heterostructure shown in Fig. 1. The length of the ferromagnet is  $d_F = 2\xi_0$ , where  $\xi_0 = \sqrt{D/\Delta}$  is the coherence length in the ferromagnet, and  $\Delta$  is the gap in the superconductors. The dotted curve shows the behavior for a normal metal between two superconductors, showing the well known minigap. The effect of a domain wall in the center of the ferromagnet is shown for several wall thicknesses  $d_W$  in Fig. 5. The new structures in the presence of the domain wall in Fig. 5 are due to spin-triplet correlations which mix between all three spin-triplet channels due to the continuous variation of the spin quantization axis across the domain wall [16]. This leads to the appearance of additional Andreev bound states inside the minigap, modifying the total density of states. In general, the presence of the domain wall reduces the

minigap, and for thick enough domain walls the minigap closes completely. This suggests that the width of the domain walls can influence the transport properties of such devices considerably.

## 6 Conclusions

We have developed a framework for studying heterostructures with superconducting and ferromagnetic parts, and have applied it to two cases: a ferromagnet in the ballistic regime with strong spin polarization ( $J \sim E_F$ ); and a weak ferromagnet ( $J \ll E_F$ ) in the diffusive regime, having an inhomogeneous, non-collinear spin magnetization (Bloch domain wall). In both cases we have studied the superconducting proximity effect with the ferromagnet sandwiched between two superconductors. We have shown, that singlet-triplet mixing occurs near interfaces between superconductors and strong ferromagnets, or near domain-wall structures in the case of weak ferromagnets in proximity to a superconductor. The origin of equal-spin triplet correlations in the two studied cases is very different. In a strong ferromagnet between two superconductors, the singlet triplet mixing takes place within the superconductor, in a layer of the order of the coherence length near the interface, enabling triplet correlations to penetrate into the ferromagnet. In contrast, in weak ferromagnets with domain walls of Bloch type, equal-spin triplet correlations are produced within the ferromagnet, near the domain walls.

## Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft within the Center for Functional Nanostructures.

## References

1. G. Eilenberger, Z. Phys. **214**, 195 (1968).
2. A. I. Larkin and Y. N. Ovchinnikov, Sov. Phys. JETP **28**, 1200 (1969).
3. J. W. Serene and D. Rainer, Phys. Rep. **101**, 221 (1983).
4. J. C. Cuevas, A. Martín-Rodero, and A. Levy Yeyati, Phys. Rev. B **54**, 7366 (1996); J. C. Cuevas and M. Fogelström, Phys. Rev. B **64**, 104502 (2001).
5. M. Eschrig, J. Kopu, J. C. Cuevas, and Gerd Schön, Phys. Rev. Lett. **90**, 137003 (2003).
6. J. Kopu, M. Eschrig, J. C. Cuevas, and M. Fogelström, Phys. Rev. B **69**, 094501 (2004).
7. Y. Nagato, K. Nagai and J. Hara, J. Low Temp. Phys. **93**, 33 (1993); S. Higashitani and K. Nagai, J. Phys. Soc. Jpn. **64**, 549 (1995).
8. N. Schopohl and K. Maki, Phys. Rev. B **52**, 490 (1995).
9. M. Eschrig, Phys. Rev. B, **61**, 9061 (2000).
10. T. Tokuyasu, J.A. Sauls, and D. Rainer, Phys. Rev. B **38**, 8823 (1988).
11. M. Fogelström, Phys. Rev. B **62**, 11812 (2000).

12. Yu.S. Barash and I.V. Bobkova, Phys. Rev. B **65**, 144502 (2002).
13. K. D. Usadel, Phys. Rev. Lett. **25**, 507 (1977).
14. Yu. V. Nazarov, Superlatt. and Microstruct. **25**, 121 (1999).
15. N. M. Chtchelkatchev and I. S. Burmistrov, Phys. Rev. B **68**, 140501(R) (2003).
16. F. S. Bergeret, A. F. Volkov, and K. B. Efetov, Phys. Rev. Lett. B **86**, 4096 (2001); R. Mélin and S. Peysson, Phys. Rev. B **68**, 174515 (2003).

# Neutron Scattering Investigations of High $T_c$ Compounds

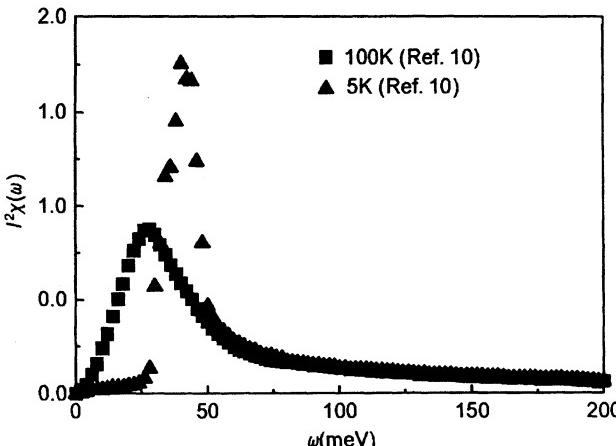
Lothar Pintschovius

Forschungszentrum Karlsruhe, IFP  
P.O.B. 3640, 76021 Karlsruhe, Germany  
[pini@ifp.fzk.de](mailto:pini@ifp.fzk.de)

**Abstract.** The high  $T_c$  compounds have been very extensively studied by neutron scattering. Whereas the properties of the undoped parent compounds are now well understood, the doping induced changes are a matter of ongoing interest. In particular, the spin fluctuations have been studied in great detail because they are widely considered to mediate high  $T_c$  superconductivity. Nevertheless, the experimental characterization of the magnetic excitations is still very incomplete for important materials. Very recent progress indicates a close similarity between the magnetic response of all hole doped compounds which will help to sort out many theoretical proposals. Phonon studies are smaller in number because phonons were often considered as irrelevant for high  $T_c$  superconductivity. Indeed, phonons were found to change rather little upon doping, with the exception of the plane polarized Cu-O bond-stretching vibrations. The anomalous dispersion of these modes and the large phonon linewidths indicate a strong electron-phonon coupling.

## 1 Introduction

Neutron scattering techniques have been widely used to explore and characterize the properties of high Tc compounds: the crystal structure and the structural phase transitions, the magnetic structure and the phase diagrams, collective magnetic excitations of 4f ions, flux lattice structures and flux lattice melting, phononic and magnetic excitations of the Cu spins [1]. Over the years, the number of neutron studies on high Tc's has declined in many areas, simply because the issues are settled now. On the other hand, there is a strong ongoing interest in the lattice vibrations and the spin excitations. It is hoped to elucidate the still unknown mechanism of high Tc superconductivity or, at least, to shed some light on the electronic ground state which theory predicts to be quite different from that of normal metals. Unfortunately, neutron studies of the phonons and of the magnetic excitations require single crystals of considerable size, i.e. of the order of one to several  $cm^3$ , which has hampered such studies for a long time. Continuous progress in crystal growth and in neutron instrumentation allowed one to obtain new results of considerable importance up to the present day. In the following, I will comment on the actual state of this particular field.

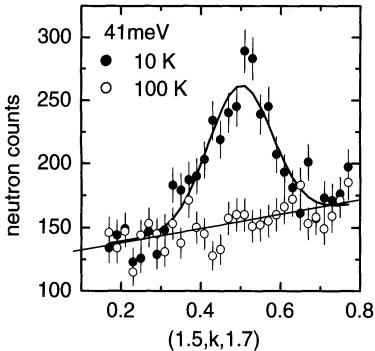


**Fig. 1.** Spin-polarized inelastic neutron scattering spectra for  $\text{YBa}_2\text{Cu}_3\text{O}_{6.92}$ . The figure was taken from Carbotte, Schachinger and Basov [2]

## 2 Magnetic Excitations

An enormous amount of neutron beam time has been devoted to inelastic neutron scattering studies of the magnetic excitations because of the widely held belief that high  $T_c$  superconductivity is mediated by spin excitations rather than by phonons. To underpin this belief, it is necessary to characterize the spin fluctuation spectrum in the normal state and in addition, to get some information on the coupling strength of the spin fluctuations to the quasiparticles. Both ingredients are necessary for an estimate of the spin fluctuation mediated  $T_c$ . Unfortunately, already the first of the two tasks has been exceedingly difficult to accomplish, both because of the weakness of the magnetic signal and because of the large range in energy involved: the spin excitations in the cuprate superconductors appear to extend over the same energy range as in the insulating parent compounds (200 meV). In fact, the state of the field is considerably less advanced than has been suggested in publications of high impact. For instance, it was disseminated in a paper by Carbotte and co-workers [2] that precise polarized neutron-scattering are now available for the spin fluctuation spectrum of a variety of the copper-oxides. Actual data were shown for  $\text{YBa}_2\text{Cu}_3\text{O}_{6.92}$  (abbr. YBCO,  $x = 0.92$ , Fig. 1).

Indeed, the figure suggests that high quality data are available for this compound extending over a very large energy range for both the superconducting and for the normal state. Such nice data, if they were real, would be a very valuable input for theories (and were used as such). Unfortunately, the reality is not that bright. To begin with, none of the data shown in this figure were obtained with polarized neutrons [3]. Using polarized neutrons is an elegant way to separate the magnetic signal from the background but in fact, only a few polarized neutron scattering studies on cuprates have been

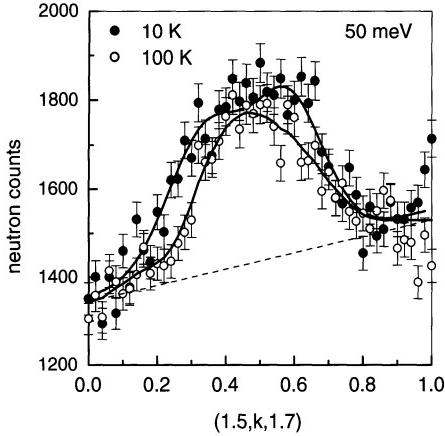


**Fig. 2.** Inelastic neutron scattering spectra taken on optimally doped YBCO along the line  $Q = (1.5, k, 1.7)$  at  $T = 10$  K (full dots) and at  $T = 100$  K (open symbols) after [9]

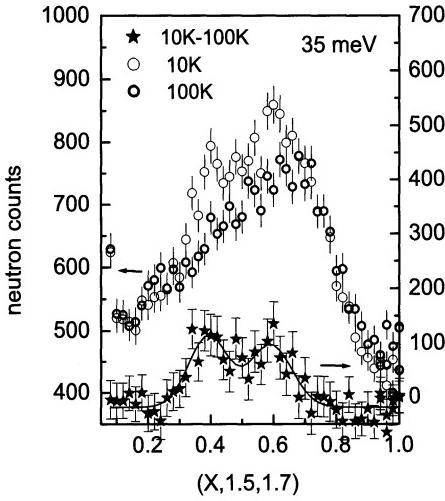
published so far. The reason is that a heavy penalty has to be paid for using polarized neutrons, i.e. an about 20fold reduction in intensity. Further, the symbols used in Fig. 1 are suggestive of data points but depict only lines drawn as a guide to eye through the raw data. What is more, the actual data (which were taken from [4]) do not go beyond  $E = 52$  meV, the rest shown in Fig. 1 seems to have been extrapolated with the help of a theory. Finally, the temperature dependence reported by Bourges et al. [4] for  $x = 0.92$  is in conflict with observations of other authors. In particular, there is strong evidence that the magnetic intensity at  $E = 41$  meV, the so-called resonance energy, drops to zero above the superconducting transition temperature for doping levels close to the optimum  $T_c$  [5,6,7,8,9]. This fact – which stirred great interest in this phenomenon – is demonstrated in Fig. 2 for a sample with  $x = 0.95$  [9]. Further, the redistribution of spectral weight from energies 10 to 30 meV to higher energies on entering the superconducting state as suggested in Fig. 1 was not seen in other studies of [6,7,8] studies of optimally or slightly overdoped YBCO and also not in a polarized neutron scattering study on underdoped ( $x = 0.7$ ) YBCO [11].

In fact, the spin fluctuation spectrum of optimally doped YBCO in the normal state is still very poorly known. A very recent study [10] has brought some progress but only for energies above the resonance energy (some raw data are shown in Fig. 3). For energies below the resonance energy, polarized neutrons seem to be indispensable to separate the magnetic signal (if there is one) from the structured background due to phonons (Fig. 4). On cooling below the superconducting transition temperature, the intensity of the magnetic scattering was found to increase at all energies accessible in the experiment. It remains to be understood where the spectral weight is drawn from.

The 10 K data depicted in Figs. 3,4 show clear evidence of incommensurate spin fluctuations at energies above and below the resonance energy. Their dispersion is shown in Fig. 5(left) [8]. The low frequency and the high frequency branches merge at the resonance peak. As is pointed out in [8], the  $q$ -integrated intensity does not vary much in the energy interval studied so



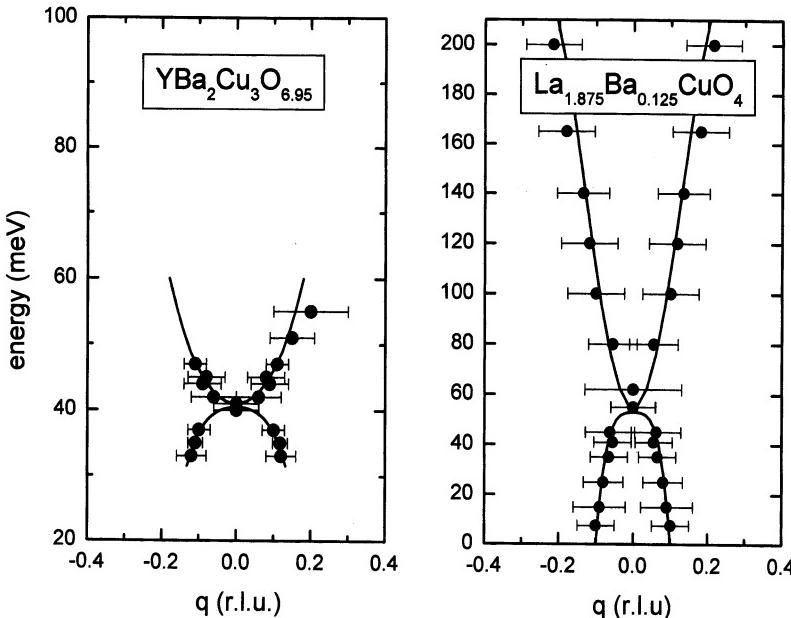
**Fig. 3.** Inelastic neutron scattering spectra taken on optimally doped YBCO at an energy above the resonance energy. The lines are a guide to the eye. [10]



**Fig. 4.** Inelastic neutron scattering spectra taken on optimally doped YBCO at an energy below the resonance energy. The line is a fit with two Gaussians. [9]

far. In contrast to what has often been suggested, the resonance peak does not dominate the magnetic excitation spectrum but stands out only because at this energy, the intensity is confined to a rather narrow area in  $q$ -space.

The theoretical framework for an understanding of the experimental results remains highly controversial. As an example, M. R. Norman proposed in 2000 a Fermi liquid-like theory for relating the incommensurability observed in neutron scattering experiments to the electronic structure [12]. In fact, the theory gives a fairly good description of the hour-glass shaped dispersion reported by Arai and co-workers for underdoped YBCO with  $x = 6.7$  [13] (which is not much different from that depicted for optimally doped YBCO in Fig. 5). A year later, the same author uses a totally different approach for explaining the observation of incommensurate spin fluctuations in YBCO [14]: they are now termed a magnetic collective mode predicted to arise below



**Fig. 5.** Left: dispersion of the spin fluctuations observed in optimally doped YBCO at  $T = 10$  K, i.e. in the superconducting state [9]. The lines are calculated from a model fitted to the data. Note that energies above 56 meV were not accessible in this experiment and that no magnetic excitations were observed below 33 meV. right: dispersion of the spin fluctuations observed in the static stripe phase  $\text{La}_{1.875}\text{Ba}_{0.125}\text{CuO}_4$  by Tranquada et al. [22]. The lines are polynomials fitted to the data

the particle-hole spin-flip continuum in the d-wave superconducting state (a very similar interpretation has been proposed by other authors [15]). Regrettably, the author does not comment on how far his new theory makes his previous proposal obsolete or might be reconciled with it. Presumably, the new approach was motivated by new data on the spin excitation spectrum in slightly underdoped YBCO ( $x = 6.85$ ) [16]. This paper reports observation of a downward dispersing branch emanating from the resonance peak but not of an upward dispersing one and also the new theory does not account for an upward dispersing branch. However, the experimental paper [16] does report observation of magnetic excitations above the resonance energy with an energy dependent width reminiscent of the dispersion seen in more underdoped [13] and in optimally doped [9,8] YBCO. These excitations are completely ignored in the second theoretical paper [14].

A widely discussed theoretical concept is that of a segregation of the holes doped into the Cu-O planes to form an array of metallic stripes acting as antiphase boundaries of antiferromagnetically ordered domains [17]. This concept was developed after the observation of incommensurate low frequency spin excitations in  $(\text{La},\text{Sr})_2\text{CuO}_4$  (LSCO) [18]. This picture gained

great popularity after the observation of magnetic and nuclear superlattice reflections consistent with static stripe order in  $\text{La}_{1.48}\text{Nd}_{0.4}\text{Sr}_{0.12}\text{CuO}_4$  by Tranquada et al. [19]. This particular compound is not – or only marginally – superconducting because static stripe order seems to compete with superconductivity. The proponents of the stripe model assume nevertheless that stripe order is a universal phenomenon in the cuprates. They argue that stripe order is dynamic in nature in the superconducting compounds. Here, I do not want to discuss the question whether or not stripe order is relevant for superconductivity -which is an open issue- but simply whether there is solid evidence for stripe order in other compounds than in  $\text{La}_{1.48}\text{Nd}_{0.4}\text{Sr}_{0.12}\text{CuO}_4$ . I note that incommensurate low frequency spin fluctuations consistent with dynamic stripe order are universally observed in  $(\text{La},\text{Sr},\text{Ba})_2\text{CuO}_4$  [20]. On the other hand, the situation is very controversial for the other well studied class of materials, i.e. YBCO. It was advocated [21] that the incommensurate spin fluctuations in YBCO discussed above are a direct analogue of those observed in LSCO. Others argue [16] that the dispersive nature of the incommensurate spin excitations speak against the stripe picture for YBCO.

I think that recent data reported by Tranquada et al. [22] for the static stripe phase  $\text{La}_{1.875}\text{Ba}_{0.125}\text{CuO}_4$  (LBCO, not superconducting) have shed new light on the situation. Using a modern spectrometer at the ISIS spallation source and a very large sample, the authors were able to obtain high resolution data over a very wide energy range (0-200 meV). At low energies, the data show the well-known intensity rods starting from the incommensurate magnetic superlattice reflections. Above 20 meV, however, the magnetic peaks show dispersion which is, moreover, very similar to that observed in YBCO (Fig. 5; right). These observations suggest a number of conclusions:

(i) an hour-glass shaped dispersion seems to be a generic feature of the hole-doped compounds. To my opinion, this observation strongly favors Fermi liquid-like theories like the one mentioned above [12] over other more exotic proposals because only Fermi liquid-like theories were able to predict both the upward and the downward dispersing branches.

(ii) the features seen in the magnetic excitation spectrum of LBCO around  $E = 50$  meV are strongly reminiscent of the resonance peak observed so far only in YBCO and in Tl- and Bi-based compounds with high  $T_c$  [23,24]. This can hardly be a coincidence and means that superconductivity is not a necessary ingredient for explaining the resonance peak as many theories have suggested. Of course, superconductivity does come into play with regard to the observed changes in intensity of the spin excitations but apparently, the opening of the superconducting gap just restores a phenomenon which is seen already in a non-superconducting cuprate. In this context, I would like to add that the very pronounced intensity changes of the resonance peak on entering the superconducting state obvious from Fig. 2 are somewhat exceptional because restricted to optimally or overdoped YBCO. More gradual changes on cooling from room temperature were found for underdoped YBCO [16,25]

and also for fully doped YBCO containing nonmagnetic impurities [26]. I note that qualitatively similar changes were observed also in LSCO [27]. The fact that the intensity changes occur in LSCO at relatively low energies (3-15 meV) is presumably a consequence of the much smaller superconducting gap.

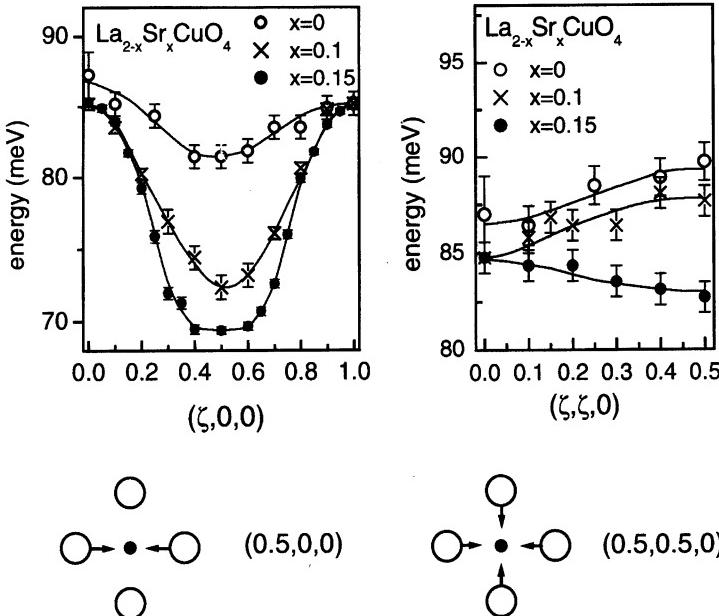
(iii) somewhat surprisingly, the magnetic excitations in the static stripe phase cannot be explained by a spin wave model, i.e. spin waves dispersing out of the magnetic superlattice peaks in the form of expanding cones [22]. Therefore, arguments against the stripe picture for YBCO based on naive expectations from a spin wave model have become pointless. On the other hand, there is no valid model relating the dispersive high frequency spin excitations to stripe order. On these grounds, the applicability of the stripe picture to YBCO remains undecided.

In summary, the neutron experiments suggest a unified picture for the magnetic response of the hole-doped compounds, at least for the basic dispersion of the spin excitations (the electron-doped compounds, however, seem not to fit into this picture [28]). There is strong evidence that the magnetic excitation spectrum extends over a very wide energy range in all cuprates. This means that it makes little sense to focus on just a particular phenomenon like the resonance peak to explain features seen by other spectroscopic methods: there is too little phase space involved. Despite all the experimental efforts undertaken so far, the normal state spin fluctuations are well characterized only for compounds with relatively low  $T_c$  like LSCO and underdoped YBCO [25]. I note that this persisting hole in our knowledge is largely due to a very simple reason: the total spectral weight of the spin fluctuations appears to decrease strongly on doping [29,25], a fact which is often forgotten by the proponents of spin fluctuation-mediated superconductivity.

### 3 Lattice Vibrations

Extensive inelastic neutron scattering measurements on a number of doped and undoped cuprates [30] revealed the following results: (i) the lattice vibrations of the undoped parent compounds can be very well understood in the framework of phenomenological models, in particular the shell model [31], developed for ionic insulators. (ii) the same models provide a very decent description even for the superconducting compounds after allowance is made for screening by free carriers [31], i.e. for the insulator to metal transition. (iii) the models just mentioned fail, however, to reproduce the strong frequency renormalization observed upon doping for in-plane Cu-O bond-stretching vibrations. More precisely, only longitudinal phonon branches show such a renormalization and zone boundary modes react much more strongly than zone center modes. Besides, stronger effects are seen in the (100) direction than in the (110) direction. An example is shown in Fig. 6 for the case of LSCO.

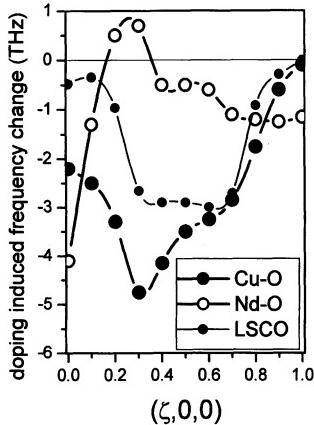
Similar effects are seen in all other compounds investigated so far. To illustrate this finding, the doping induced changes in the plane-polarized



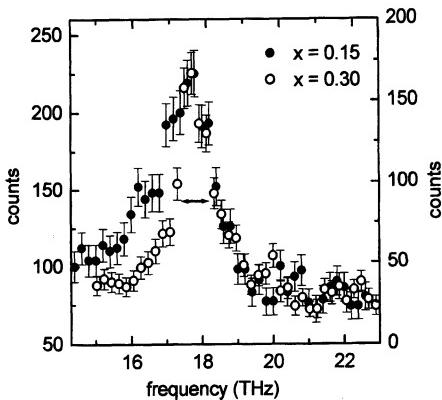
**Fig. 6.** Dispersion of the longitudinal in-plane Cu-O bond-stretching vibrations in undoped, underdoped and optimally doped LSCO [30]. The displacement patterns for the zone boundary modes are shown as well. The mode for  $q = (0.5, 0.5, 0)$  is the planar breathing mode. The mode for  $q = (0.5, 0, 0)$  is often termed "half-breathing" mode. Note that the zone boundary in the (100) direction is at  $(0.5, 0, 0)$  and that the point  $(1, 0, 0)$  is the Z-point

bond-stretching branches are depicted in Fig. 7 for the electron-doped compound  $Nd_{1.85}Ce_{0.15}CuO_4$  (NCCO) [32]. NCCO has a somewhat different structure than LSCO and therefore, there are two types of plane-polarized bond-stretching branches: one type with elongations in the Cu-O planes and the other with elongations in the Nd-O planes. Both types of vibrations are strongly renormalized at the zone center which is due to screening by the free carriers. However, the screening effects decrease rapidly when going away from the zone center which means that NCCO is a poor metal. As can be seen in Fig. 7, only the Cu-O vibrations are strongly renormalized upon doping for  $q > 0.1$  r.l.u.. Concurrently, only Cu-O modes show strong line broadenings [32] similarly to what has been observed in optimally doped LSCO [33].

Recently, it has been reported [34] that the bond-stretching modes in overdoped LSCO ( $x = 0.3$ , not superconducting) behave like those in the undoped parent compound indicating that the phonon softening is correlated with  $T_c$ . Inelastic neutron scattering experiments on similar samples [35] did NOT confirm this result. Although the phonons harden slightly when going from optimally doped to overdoped LSCO around  $q = (0.3, 0, 0)$  (Fig. 8),



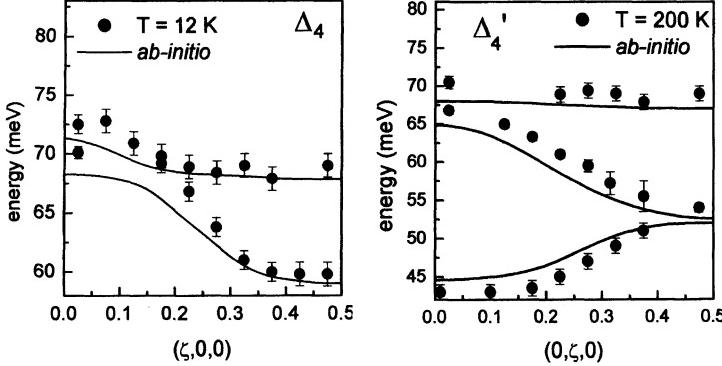
**Fig. 7.** Doping induced frequency changes in the plane-polarized Nd-O and Cu-O bond-stretching vibrations in optimally doped NCCO [32]. For comparison, the corresponding changes in optimally doped LSCO [33] are shown as well



**Fig. 8.** Inelastic neutron scattering data taken for  $q = (0.3, 0, 0)$  on optimally doped (left scale) and on overdoped (right scale) LSCO [35]

the zone boundary mode is even slightly softer in overdoped LSCO. On the other hand, the strong reduction in linewidth upon overdoping in the (100) direction for  $0.1 < q < 0.4$  (Fig. 8) indicates a significant reduction of the electron-phonon coupling of these modes.

The experimentally determined phonon dispersion in optimally doped YBCO is very well reproduced by density functional theory based on the local density approximation (LDA) [36]. I emphasize that the anomalous downward dispersion of the bond-stretching branches [38] is also correctly described by this theory (Fig. 9). Somewhat surprisingly, this theory predicts rather small linewidths for these modes [37], i.e. only a very moderate electron-phonon coupling strength. As a consequence, the experimentally observed linewidths are seriously underestimated (by a factor 5-10). Nevertheless, it is unlikely that this discrepancy invalidates the basic conclusion of Bohnen et al. [36] that conventional electron-phonon coupling is too weak in YBCO to explain the high  $T_c$ : the contribution of the few phonon branches in question to the total electron-phonon coupling strength is not large enough to warrant such a conclusion.

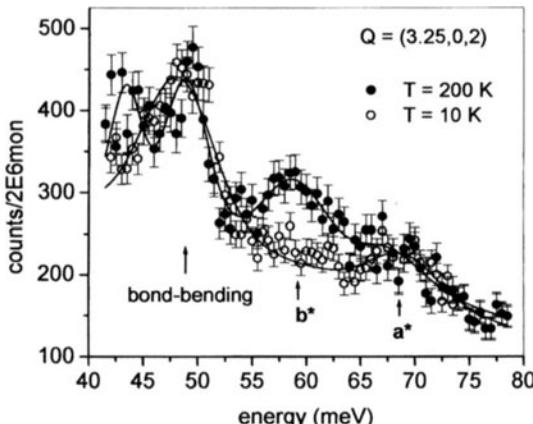


**Fig. 9.** Dispersion of oxygen phonon branches in optimally doped YBCO in the (100) and in the (010) direction, respectively. The flat branches around  $E = 68$  meV are c-polarized apical O bond-stretching vibrations whereas the downward dispersive branches are in-plane polarized Cu-O bond-stretching modes. The upward dispersing branch in the right hand panel has bond-bending character [9]. The *ab-initio* results were taken from [36]

There is another shortcoming of the LDA calculations: they are in very good agreement with high temperature data for the (010) direction (Fig. 10) but not with low temperature data (for which the calculations are actually made). As is demonstrated in Fig. 10, a pronounced downward shift of spectral weight was observed upon cooling within a narrow range of wave vectors in the (010) direction (no such effect was seen in the (100) direction). As has been outlined in [38], the temperature evolution indicates that optimally doped YBCO is close to a charge density wave instability. At present, it remains unclear whether this instability is driven by Fermi surface nesting or by correlation effects as predicted by theories for stripe order [17]. I emphasize that LDA calculations automatically account for nesting effects. This speaks against Fermi surface nesting as the driving force of the phonon anomaly unless the calculated Fermi surface is somewhat unrealistic due to improper treatment of the correlations. In any case, one has to go beyond the LDA to understand the instability.

#### 4 Concluding Remarks

As outlined above, there is a solid basis for the understanding of the lattice vibrations in the high  $T_c$  compounds, both experimentally and theoretically, although some questions like the origin of the charge density wave instability in YBCO remain to be answered. As for the spin fluctuations, our understanding is much less advanced despite the tremendous experimental efforts already undertaken and the very many theoretical papers devoted to this issue. On the experimental side, there is substantial progress to this very day giving rise to hopes that we will soon have a relatively clear picture of the spin



**Fig. 10.** Energy scans taken on a twinned sample of optimally doped YBCO at a nominal wave vector of  $Q = (3.25, 0, 2)$ . The peaks observed at  $E = 68$  meV and 59 meV are assigned to the in-plane Cu-O bond-stretching vibrations along the  $a$ - and the  $b$ -direction, respectively, whereas the 49 meV peak is related to Cu-O bond-bending vibrations. On cooling, the 59 meV peak is depleted which is partly compensated by a gain in intensity around 45 meV. Some of the missing intensity is presumably shifted down to energies below 40 meV [38]

fluctuations in some key compounds. This will finally provide a suitable basis for the theoretical understanding of the spin fluctuations and their relevance for high  $T_c$  superconductivity.

## References

1. For a review see: Neutron Scattering in Layered Copper-Oxide Superconductors, in *Phys. and Chem. of Materials with Low-Dimensional Structures*, Vol. 20, A. Furrer (Ed.) (Springer, Berlin, Heidelberg 1996).
2. J. P. Carbotte, E. Schachinger and D. N. Basov, Nature **401**, 354 (1999).
3. The data shown in [2] were taken from P. Bourges et al. [4]. The authors of this paper do not claim to have used polarized neutrons.
4. P. Bourges et al., cond-mat/9902067.
5. H. F. Fong et al. Phys. Rev. Lett. **75**, 316 (1995).
6. P. Bourges et al., Phys. Rev. B **53**, 876 (1996).
7. P. Dai et al., Phys. Rev. B **63**, 054525 (2001).
8. D. Reznik, P. Bourges, L. Pintschovius, Y. Endoh, Y. Sidis, Y. Shiokara, S. Tajima, cond-mat/0307591.
9. L. Pintschovius et al.: to appear in Proc. of ISS 2003, October 2003 at Tsukuba, Japan, to be published as a special issue of Physica C.
10. D. Reznik, L. Pintschovius, Y. Endoh, S. Tajima, unpublished results Pintschovius, Y. Endoh, Y. Sidis, Y. Shiokara, S. Tajima, cond-mat/0307591.
11. H. F. Fong et al. Phys. Rev. Lett. **78**, 713 (1997).
12. M. R. Norman, Phys. Rev. B **61**, 14751 (2000).

13. M. Arai et al., Phys. Rev. Lett. **83**, 608 (1999).
14. M. R. Norman, Phys. Rev. B **63**, 092509 (2001).
15. F. Onufievra and P. Pfeuty, Phys. Rev. B **65**, 054515 (2002).
16. P. Bourges et al., Science **288**, 1234 (2000).
17. J. Zaanen and O. Gunnarsson, Phys. Rev. B **40**, 7391 (1989).
18. G. Shirane et al., Phys. Rev. Lett. **63**, 330 (1989).
19. J. M. Tranquada et al., Nature **375**, 561 (1995).
20. see, e.g., K. Yamada et al., Phys. Rev. B **57**, 6165 (1998).
21. H. A. Mook et al., Nature **395**, 580 (1998).
22. J. M. Tranquada et al., cond-mat/0401621
23. H. He et al., Phys. Rev. Lett. **86**, 1610 (2001).
24. H. He et al., cond-mat/0201252.
25. H. A. Fong et al., Phys. Rev. B **61**, 14773 (2000).
26. H. A. Fong et al., Phys. Rev. Lett. **82**, 1939 (1999).
27. T. E. Mason et al., Phys. Rev. Lett. **77**, 1604 (1996).
28. K. Yamada et al., Phys. Rev. Lett. **90**, 137004 (2003).
29. L.P. Regnault, Ph. Bourges and P. Burlet: Phase diagrams and spin correlations, in *Phys. and Chem. of Materials with Low-Dimensional Structures*, Vol. 20, A. Furrer (Ed.) (Springer, Berlin, Heidelberg 1996) pp. 85-134.
30. L.Pintschovius and W. Reichardt: Phonon dispersions and phonon density-of-states, in *Phys. and Chem. of Materials with Low-Dimensional Structures*, Vol. 20, A. Furrer (Ed.) (Springer, Berlin, Heidelberg 1996) pp. 85-134.
31. S. L. Chaplot, W. Reichardt, L. Pintschovius and N. Pyka, Phys. Rev. B **52**, 7230 (1995).
32. M. Braden, L. Pintschovius, K. Yamada, D. Reznik, unpublished results.
33. L. Pintschovius and M. Braden, Phys. Rev. B **60**, R15039 (1999).
34. T. Fukuda et al., cond-mat/0306190.
35. L. Pintschovius, K. Yamada, D. Reznik, unpublished results.
36. K. P. Bohnen, R. Heid and K. Krauss, Europhys. Lett. **64**, 104 (2003).
37. R. Heid, private communication.
38. L. Pintschovius et al., cond-mat/030837.

# Index

- ab-initio calculation, 415  
absorber material, 27  
Abstreiter, G., 213  
adiabatic passage, 181  
AFM, 327  
ageing, 389  
– definition, 390  
– regime, 391  
AlAs, 339  
Aleiner, I.L., 147  
Alet, F., 265  
amorphization, 339  
amorphous material, 339  
angle-resolved photoelectron spec-  
troscopy, ARPES, 241  
anomalous refraction, 105  
antiferromagnetic, 255  
Apalkov, V., 81  
atomic force microscopy, 327  
atomic wire, 277  
Augustin, M., 117  
Aumaître, S., 401  
  
Baldassarri, Höger von Högersthal, G.,  
    191  
ballistic limit, 533  
band structure, 93  
Barth, S., 363  
Batrouni, G., 265  
Bauer, G., 227  
Bayer, M., 191  
Beale, T.A.W., 467  
Bencok, P., 467  
Biberacher, W., 213  
Bichler, M., 81  
binary agent resource, 427  
Bläsing, J., 313  
Blanco, A., 93  
  
bond-stretching vibration, 547  
Boothroyd, A.T., 467  
Born approximation  
– self consistent, 169  
Borri, P., 191  
Bose-Einstein condensation, 277  
Boson  
– Mott-insulator transition, 277  
bosonic atom  
– ultra-cold, 265  
Bragg reflection, 227  
Brandes, T., 181  
Brandt, M.S., 453  
Breu, A.P.J., 401  
Brieler, F.J., 491  
Broocks, K.-B., 81  
Busch, K., 93  
Busch, P., 327  
  
causality, 397  
CdS, 27  
CdTe, 13, 51  
CeCoIn<sub>5</sub>, 260  
Chakraborty, T., 81  
charge-transfer salt, 241  
charged defect, 503, 504  
Chen, L., 491  
Choe, S.C., 427  
Christen, J., 313  
Claessen, R., 241  
classical liquid, 439  
Clos, R., 313  
cluster expansion method, 415  
Co, 479  
COBE experiment, 382  
coherence length  
– superconducting, 533  
complex system, 427

- computational materials science, 415
- conductance peak, 213
- conformal group, 397
- conversion efficiency, 51
- copolymer, 327
- Corbino ring, 135
- correlation, 277
  - singlet, triplet, 533
- correlation effects, 241
- correlator, 439
- cosmic microwave background
  - radiation, 375
- cosmic structure, 375
- cosmology, 375
- Coulomb drag, 169
- Coulomb interaction, 241, 503, 504
- critical slowing down, 265
- crowding, 427
- crystal growth, 299
- $\text{Cu}(\text{In},\text{Ga})(\text{S},\text{Se})_2$ , 13
- $\text{Cu}(\text{In},\text{Ga})\text{Se}_2$ , 51
- $\text{Cu}(\text{In},\text{Ga})(\text{Se},\text{S})_2$ , 27
- $\text{Cu}(\text{In},\text{Ga})\text{Se}_2$ , 27
- $\text{Cu}(\text{In},\text{Ga},\text{Al})(\text{Se},\text{S})_2$ , 27
- Cuevas, J.C., 533
- $\text{CuGaSe}_2$ , 27
- $\text{CuInSe}_2$ , 27
- current domain, 147
- Dadgar, A., 313
- Daumiller, I., 313
- defect cluster, 505
- degradation, 39
- demixing of grains, 401
- density functional theory, 415
- density of states, 169
- detailed balance, 3
- Deubel, M., 93
- DFT, 415
- dielectric constant, 69
- dielectric response, 69
- Diez, A., 313
- Dmitriev, I.A., 147
- domain wall, 533
  - depinning, 479
  - propagation, 479
- double-layer system, 169
- dynamical
  - exponent, 390
  - scaling, 389
  - symmetry, 389
- dynamical scaling, 390
- efficiency, 3
- Eilenberger-Larkin-Ovchinnikov transport equation, 533
- El Farol Problem, 427
- electron plasma, 69
- electron-phonon coupling, 241, 547
- electronic correlation, 503
- energy resolved mass spectrometry, 299
- Enkrich, C., 93
- entanglement, 191, 439
- Eschrig, M., 533
- evanescent mode, 69
- Ewald sphere, 467
- exciton
  - charged, 81, 191
  - dephasing, 191
  - photoluminescence, 191
- excluded volume, 439
- exponent
  - autocorrelation, 391, 399
  - autoresponse, 391, 399
- Fafard, S., 191
- Fermi liquid behavior, 241
- ferromagnet, 479
- ferromagnetic semiconductor, 515
- financial market, 363, 427
- fluorescence imaging, 69
- flux lattice melting, 547
- Fogelström, M., 533
- Forchel, A., 191
- formation enthalpy, 415
- four-wave mixing, 191
- Fröba, M., 491
- fractal exponent, 378
- fractional quantum Hall effect, 81
- Friedrich, R., 363
- function
  - autocorrelation, 390
  - autoresponse, 390
  - correlation, 389, 399
  - response, 389
- Furdyna, J.K., 515
- $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ , 453

- (Ga,Mn)As, 503
- Götze, 439
- Ga<sub>1-x</sub>Mn<sub>x</sub>As, 515
- GaAs, 339
  - Mn-doped, 503
  - GaAs/AlGaAs, 81, 135, 147
  - galaxy
    - distribution of, 375
    - structure, 375
  - Galilei invariance, 392, 396, 398
  - GaN, 313, 339
    - devices, 313
    - epitaxial growth, 313
    - heteroepitaxy, 313
    - molecular beam epitaxy (MBE), 313
  - GaP, 339
  - García-Vidal, F. J., 69
  - geometry
    - fractal, 377
  - GISAXS, 327
  - glassy dynamics, 439
  - Glauber-Ising model, 389, 390, 394, 396
  - Goennenwein, S.T.B, 453
  - Goetzendorfer, A., 401
  - Gornyi, I.V., 169
  - Gourley, S., 427
  - granular system, 401
  - gravitational clustering, 383
  - Grayson, M., 213
  - grazing incidence x-ray reflectometry (GIXR), 299
  - Grochowski, R., 401
  - Gross, R., 453
  - ground state
    - electronic, 547
  - Hänggi, P., 157
  - Hall bar, 135
  - Hall resistance, 135
  - hard rod, 439
  - harmonic trap, 265
  - Hatton, P.D., 467
  - Hawrylak, P., 191
  - heavy fermion metal, 253
  - Heimbrodt, W., 491
  - Heitmann, D., 81
  - Hempel, T., 313
  - Henkel, M., 389
  - Hermatschweiler, M., 93
  - Hesse, A., 227
  - heterojunction, 81
    - acceptor-doped, 81
  - heterostructure, 135, 147
    - ballistic, 533
  - Heyn, C., 81
  - Hezel, R., 39
  - high  $T_c$  compound, 547
  - high  $T_c$  superconductivity, 547
  - high refractive index material, 93
  - Hippler, R., 299
  - hole doped compound, 547
  - holographic lithography, 93
  - Holy, V., 227
  - Hubbard model, 241
    - bosonic, 265
  - Huber, M., 213
  - Huebl, H., 453
  - Hui, Pak Ming, 427
  - hybrid structure
    - superconducting, ferromagnetic, 533
  - III-V compound, 339
  - Iliev, R., 117
  - impurity problem, 277
  - impurity spin, 503
  - InAs, 339
  - Indium Tin Oxide, ITO, 299
  - inelastic relaxation rate, 147
  - infra-red absorption, 453
  - InP, 339
  - InSb, 339
  - intermetallic compound, 415
  - invariance
    - conformal, 392
  - ion bombardment, 339
  - Jahn-Teller distortion, 467
  - Jarrett, T., 427
  - Johal, T., 467
  - Johnson, N.F., 427
  - Kaluza, A., 313
  - Kamp, M., 313
  - Kerr effect
    - magneto-optical, 481
  - Klar, P.J., 491
  - Klein, A., 13
  - Kläui, M., 479

- Koch, W., 93  
 Koeder, A., 453  
 Kohler, S., 157  
 Kondo effect, 253  
 Kondo lattice, 261  
 Konstandin, A., 533  
 Kopu, J., 533  
 Korkusinski, M., 191  
 Krost, A., 313  
 Krtschil, A., 313  
 Kruelle, C.A., 401  
 Krug von Nidda, H.-A., 491  
 Kunze, M., 313  
 Kurz, T., 491
- Landau level, 169  
 Langbein, W., 191  
 Larionov, A., 191  
 Lederer, F., 117  
 Lehmann, J., 157  
 light propagation, 117  
 light-emitting diode, 313  
 Lim, W.L., 515  
 Limpert, J., 117  
 Linden, S., 93  
 liquid-solid transition, 401  
 Liu, X., 515  
 local scale invariance, 389, 393  
 Loidl, A., 491  
 Luttinger-liquid, 277  
 Lyanda-Geller, Y.B., 191  
 Lévy distribution, 363
- Müller, S., 415  
 magnetic  
   – anisotropy, 515  
   – excitation, 548  
   – moment, 453  
   – nanostructure, 491  
   – ordering, 453  
 magneto-electronic device, 453  
 magneto-oscillation, 147  
 magneto-resistance, 135  
 Mani, R.G., 135  
 many body state, 277  
 Martín-Moreno, L., 69  
 Maxwells Theory, 105  
 MBE, 227  
 Meisel, D.C., 93
- metal alloy, 415  
 metal oxide film, 299  
 metal-insulator transition, 503  
 microwave radiation, 135, 147  
 Mirlin, A.D., 147, 169  
 Mn, 453, 508  
   – acceptor, 453  
 mode coupling theory, 439  
 Modlich, A., 313  
 molecular beam epitaxy, 227  
 molecular wire, 157  
 momentum-resolved magneto-tunneling, 213  
 Monte-Carlo simulation, 415  
 Mott-Hubbard-transition, 277  
 Mott-insulator, 265
- nano-structured metal film, 69  
 nanocrystal  
   – Si, 351  
 nanowire  
   – (Zn,Mn)S, 491  
 neutron scattering, 547  
 NiFe, 479  
 Nolte, S., 117  
 non-equilibrium, 135  
   – critical dynamics, 390  
   – exponents, 391  
 non-Fermi liquid electronic excitation, 253
- observation time, 390  
 Opel, M., 453  
 optical  
   – lattice, 277  
   – spectrum, 93  
   – transmission, 69  
   – waveguide, 105  
 orbital ordering, 467  
 order parameter, 253, 389, 401  
 organic conductor, 241  
 Ortner, G., 191  
 Ozin, G.A., 93
- pair correlation, 422  
 Papadakis, C.M., 327  
 paramagnon, 253  
 Peinke, J., 363  
 Pereira, S., 93

- Peschel, U., 117  
 phase transition, 253, 401  
 phase-ordering kinetics, 390  
 Philipp, J.B., 453  
 photo-excitation, 135  
 photoelectron spectroscopy, 13  
 photoluminescence, 81  
 photonic bandgap device, 105  
 photonic crystal, 105, 117  
 – band structure, 93  
 – three dimensional, 93  
 photoresistivity, 147  
 photovoltaic market, 51  
 photovoltaic material, 27  
 photovoltaics, 39  
 physical ageing, 389  
 Pietronero, L., 375  
 Pintschovius, L., 547  
 plasma frequency, 69  
 plasma-assisted deposition, 299  
 point defect, 339  
 poly(styrene-*b*-butadiene), 327  
 poly(styrene-*b*-methylmethacrylate), 327  
 Polyakov, D.G., 147  
 polymer film, 327  
 population  
 – adaptive agent, 427  
 – competitive, 427  
 – multi-agent, 427  
 Posselt, D., 327  
 power conversion efficiency, 27  
 Prabhakaran, D., 467  
 precipitation, 415  
 probability density, 363  
 properties  
 – electrical, 227  
 – optical, 227  
 protective coating, 299  
 proximity effect, 533  
 Quaas, M., 299  
 quantizing magnetic, 81  
 quantum  
 – devices, 191  
 – dot pairs, InAs/GaAs, 191  
 – dots, 191  
 – information processing, 191  
 quantum coherence, 181  
 quantum critical  
 – magnet, 253  
 – metal, 253  
 – point, 253  
 quantum critical behavior, 265  
 quantum Hall  
 – edge tunneling, 213  
 – effect  
 – integer, 135  
 – regime, 169  
 quantum Monte Carlo simulation, 265  
 quantum optics, 181  
 quantum phase transition, 277  
 quasi-classical theory, 533  
 quasihole, 81  
 Queisser, H.J., 3  
 Röwf, T., 299  
 radiative recombination, 3  
 Rau, U., 27  
 reciprocal space map, 227  
 reflection characteristic, 135  
 refractive index, 117  
 Reinecke, T.L., 191  
 Reithmaier, J.P., 191  
 Renner, C., 363  
 resistance  
 – electrical, 135  
 response function, 389, 394, 398  
 Riess, F., 363  
 ring magnet, 479  
 Rother, M., 213  
 roughness  
 – surface, 363  
 Rutherford backscattering spectrometry, 339  
 scaling, 255, 363  
 scaling function, 389  
 scanning electron micrograph, 93  
 scattering  
 – charge, 467  
 – magnetic, 467, 549  
 Schön, G., 533  
 Schüller, C., 81  
 Schülli, T., 227  
 Schilling, R., 439  
 Schmidt, O.G., 227  
 Schnautz, T., 401

- Schoch, W., 453  
 Schrödinger group, 392, 397  
 Schrödinger invariance, 392, 399  
 Schröter, P., 81  
 Schreiber, T., 117  
 Schulze, F., 313  
 Schwab, M., 191  
 sea mouse, 117  
 segregation, 401  
 self-assembled semiconductor nano-  
     structure, 227  
 semiconductor, 503  
     – ferromagnetic, 453  
     – magnetic, 503  
 Shubnikov-de Haas effect, 135  
 SI  
     – amorphous, 51  
 Si, 227, 313, 351  
 Si wafer, 51  
 Si, Q., 253  
 SiC, 313  
 Siefert, M., 363  
 Siewert, J., 181  
 SiGe, 227  
 SiGe/Si(001) island sample, 227  
 silicon, 3  
 Sing, M, 241  
 Sloan Digital Sky Survey, 379  
 slowing down, 439  
 Smilgies, D.-M., 327  
 Smoluchowski equation, 441  
 solar cell, 3  
     – crystalline Silicon, 39  
     – efficiency, 39  
     – industrial, 39  
 solid solution, 422  
 Soukoulis, C.M., 93  
 spectroscopy  
     – secondary ion mass, 453  
 spin  
     – fluctuation, 547  
     – manipulation, 191  
     – ordering, 467  
     – switching, 479  
 spin fluctuation, 549  
 spin-charge separation, 241  
 spintronics, 503, 533  
 Sputnik, 3  
 standard cosmological model, 378  
 Stangl, J., 227  
 static correlations, 439  
 statistic  
     – fat tail, 363  
     – non-Gaussian, 363  
 statistical physic, 375  
 Steffen, H., 299  
 Stoffel, M., 227  
 Strass, M., 157  
 Strassburger, G., 313  
 Stutzmann, M., 453  
 sub-wavelength optics, 69  
 superconducting transition tempera-  
     ture, 549  
 superconductor/half-  
     metal/superconductor device,  
     533  
 superfluid, 265  
 surface plasmon, 69  
 switching, 479  
 Sylos Labini, F., 375  
 system  
     – binary, 401  
     – quantum Hall, 81  
 Tünnermann, A., 117  
 Tétreault, N., 93  
 thermo-remanent magnetization, 395  
 thin film  
     – deposition, 13  
     – solar cell, 13  
 thin-film solar cell, 27  
 three-level atom, 182  
 three-point function, 399  
 Timm, C., 503  
 Tomonaga-Luttinger liquid, 241  
 transition metal oxide, 467  
 transition-metal ion, 503  
 transport properties, 503  
 Troyer, M., 265  
 TTF-TCNQ, 241  
 Tun, T.M., 299  
 tunnel conductance, 213  
 turbulence, 363  
 two-dimensional electron system, 81,  
     169  
 two-time correlation, 389

- ultrafast optics, 117  
Vavilov, M.G., 147  
Veit, P., 313  
vertex, 439  
virtual crystal approximation, 511  
Vlasko-Vlasov, V.K., 515  
von Oppen, F., 169, 503  
von Freymann, G., 93  
vortex, 480  
  
Waag, A., 453  
wafer, 39  
– Silicon, 39  
waiting time, 390  
Walzel, P., 401  
Ward identity, 392  
Wasilewski, Z., 191  
Wassner, T.A., 453  
waveguide, 117  
Wegener, M., 93  
Wegscheider, W., 81, 213  
Welp, U., 515  
Wendler, E., 339  
Werner, J.H., 51  
Wesch, W., 339  
  
Wessel, S., 265  
Wilkins, S.B., 467  
Will, M., 117  
Woggon, U., 191  
Wojtowicz, T., 515  
Wulff, H., 299  
Wächter, M., 363  
  
X-ray diffraction, 227, 467  
– soft resonant, 467  
X-ray photoelectron spectroscopy (XPS), 299  
  
Yakovlev, D.R., 191  
Yugova, I., 191  
  
Zacharias, M., 351  
Zellmer, H., 117  
Zener model, 511  
Zengerle, R., 105  
zero-field-cooled susceptibility, 395  
zero-resistance state, 135, 147  
ZnO, 13, 27  
Zwanzig-Mori formalism, 445  
Zwerger, W., 277